

SEP 18 1990

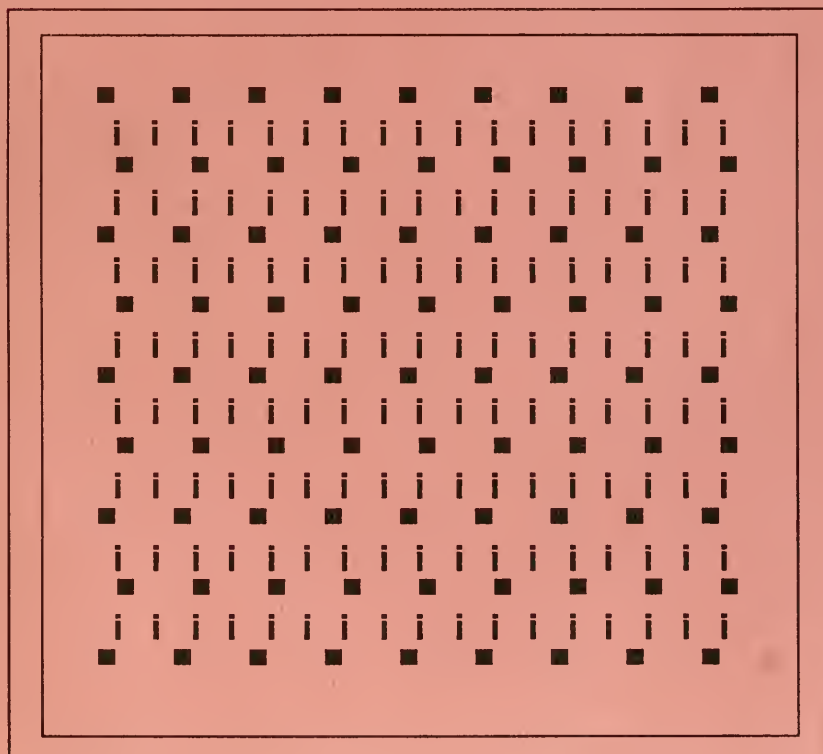
IASSIST

Q U A R T E R L Y

VOLUME 14

Spring 1990

NUMBER 1



IASSIST

QUARTERLY

The IASSIST QUARTERLY represents an international cooperative effort on the part of individuals managing, operating, or using machine-readable data archives, data libraries, and data services. The QUARTERLY reports on activities related to the production, acquisition, preservation, processing, distribution, and use of machine-readable data carried out by its members and others in the international social science community. Your contributions and suggestions for topics of interest are welcomed. The views set forth by authors of articles contained in this publication are not necessarily those of IASSIST.

Information for Authors

The QUARTERLY is published four times per year. Articles and other information should be typewritten and double-spaced. Each page of the manuscript should be numbered. The first page should contain the article title, author's name, affiliation, address to which correspondence may be sent, and telephone number. Footnotes and bibliographic citations should be consistent in style, preferably following a standard authority such as the University of Chicago press *Manual of Style* or Kate L. Turabian's *Manual for Writers*. Where appropriate, machine-readable data files should be cited with bibliographic citations consistent in style with Dodd, Sue A. "Bibliographic references for numeric social science data files: suggested guidelines". *Journal of the American Society for Information Science* 30(2):77-82, March 1979. If the contribution is an announcement of a conference, training session, or the like, the text should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event. Book notices and reviews should not exceed two double-spaced pages. Deadlines for submitting articles are six weeks before publication. Manuscripts should be sent in duplicate to the Editor: Walter Piovesan, Research Data Library, W.A.C. Bennett Library, Simon Fraser University, Burnaby, B.C., V5A 1S6 CANADA. (604) 291-4349 E-Mail: USERDLIB@SFU.BITNET. Book reviews should be submitted in duplicate to the Book Review Editor: Daniel Tsang, Main Library, University of California P.O. Box 19557, Irvine, California 92713 USA. (714) 856-4978 E-Mail: DTSANG@UCI.EDU

Title: Newsletter - International Association for Social Science Information Service and Technology
ISSN - United States: 0739-1137 Copyright 1985 by IASSIST. All rights reserved.

CONTENTS

Volume 14

Number 1

Spring 1990

FEATURES

- 4 **The Regular Labour Force Survey as a Quality Survey of the Finnish 1985 Census**
by Aarno Laihonon
- 8 **Post Censal Surveys in Great Britain**
by Robert Barnes
- 13 **The Canadian Experience with Post Censal Surveys**
by A. D. Furrie & C. McKie
- 16 **Major Post Censal Redesign of Household Sample Surveys in the United States**
by P. J. Waite
- 21 **From Census To Integrated Population Data Or Socio-Demographic Accounts**
by M. Vliegen
- 37 **The Impact of Microcomputer Technologies on Dissemination of Integrated Social, Demographic and Related Statistics**
by R. Johnston & G. Templeman
- 42 **Comparative Charting of Social Change in Four Industrialized Societies**
by Simon Langlois
- 49 **The Contemporary Jewry Database: Coordination and Flexibility in Bibliographical Registration and Retrieval**
by Mira Levine

The Regular Labour Force Survey as a Quality Survey of the Finnish 1985 Census

by Aarno Laihonon¹

Central Statistical Office of Finland

The paper deals with experiences from using the sample-based regular monthly Labour Force Survey as a quality survey of census data on the economic activity of the population. The method uses record linkage at the micro level between the data of persons in the Labour Force Survey sample and the data of the same persons in the census file. An exact linkage is made possible by the uniform personal identifier used in the Finnish population registration system.

The use of the Labour Force Survey as a reference quality survey of the census was possible because the survey week and the census week coincided, and because the survey and the census measured the same variables according to the same concepts and classifications. In this way, savings were achieved in the cost of the census quality survey. For the purpose of the quality survey, after the regular survey interview a few additional questions were asked of a sub-sample consisting of one-fifth of the Labour Force Survey sample (about 2,300 persons).

To ensure as errorless results as possible, the data of the sub-sample were reprocessed after being entered and coded as usual. The additional information on the sub-sample was utilized when estimating the final errors on the basis of a micro comparison between the census data and the survey data of the whole Labour Force Survey sample.

Introduction

In Finland, modern population and housing censuses have been carried out in 1950, 1960, 1970 and 1980, and so-called mid-decade censuses in 1975 and 1985. The latest census of 1985 was carried out within extremely tight budget constraints imposed on the Central Statistical Office by the Ministry of Finance. The tight budget constraints also affected the production of census data. New cost-saving devices had to be used.

A general outline of the 1985 census

To give an idea of the cost frame, the direct costs of the 1980 census amounted to about 80 million marks (about 17 million US dollars) and those of the previous mid-decade census of 1975 to about 26 million marks, in 1986 prices. The total expenditure of the 1985 census was not to exceed 18 million marks.

The central point of departure for the planning of the system solutions of the November 1985 census was to minimize the amount of manual work in census data collection and processing.

First of all, data collection was minimized by an extensive use of registers and administrative records when gathering the basic census data. This was largely made possible by the comprehensive, high-quality population registration, taxation and social security systems characteristic of all the Nordic countries. The use of registers and administrative records in population and housing censuses has increased steadily since the 1970 census. This development has been aided by the widespread use of the uniform personal identifier in different registers and administrative records.

A significant improvement in the register situation, which helped a great deal in bringing down the cost of the latest census, was the establishment of a building and dwelling register for Finland on the basis of the 1980 population and housing census. This register is operated in connection with the Central Population Register, and it allows the linkage of persons and dwellings. Therefore, no questions on housing were needed on the 1985 census form.

The census of November 1985 used only one questionnaire, namely for gathering employment data. The Central Population Register was used as the mailing list for the population of working age. *The questions and instructions on the census form are presented in Annex 1.*

The questionnaires were sent out by mail from the Central Statistical Office (CSO) and were returned by mail to the CSO. No local census organization was used. Another special feature of this census was that the census form was preprinted, not only with the respondent's name and address but also, for about half the population, with the name of the respondent's workplace as it appeared in the 1980 census and with the respondent's occupational title obtained from the Central Population Register. Persons obligated to respond needed only to report any changes or errors that had occurred in this information.

The census forms went directly to data entry, which was carried out as key entry. Naturally, only changes and additions on the forms had to be keyed. In this way

complete 'pictures' of the census forms were converted to machine-readable form before any other processing operations were performed. This enabled batch mode checking and correcting of the form data, leaving only about 10 per cent of the forms to be checked and corrected manually, on terminals. Next, extensive automatic coding was applied to workplace and occupation data. Thus, the number of forms requiring manual processing was drastically reduced in all phases of processing. This reduced the total cost of the census and allowed preliminary publication of the most essential census data as early as December 1986.

In the final phase of data collection, register data were also used to obtain, by imputation, the census form data of non-respondents. In this way, a satisfactory 98.6 per cent total coverage was achieved for the central data of the census form. This also contributed to the relatively small regional variation in coverage, even though the census was carried out as a direct mail-out, mail-back system without any local census organization and with only one reminder sent to non-respondents. Register imputation of questionnaire data was tried for about 139,000 persons (3.7 per cent of the population of working age), 84,000 of whom were non-respondents and the rest persons whose responses were incomplete.

The regular Labour Force Survey, the study week of which coincided with the census week, and the 1985 household survey were used as quality surveys of the census. According to the quality surveys, the general quality of the 1985 census data is significantly better than that of the previous mid-decade census of 1975. The quality of employment data is in part slightly inferior to the quality of employment data in the 1980 census.

The setup of the quality surveys of the 1985 census

Because of the high-quality, up-to-date information obtainable from the Central Population Register (CPR) in Finland, the main purpose of the population and housing census is not to count the population, but to produce data on the economic activity and housing conditions of the whole population. The resident population of the country as registered in the CPR was the population of the census.

Thus, from the point of view of the quality of the census data, there was not, by definition, any undercount or overcount of the population. The aim of the census quality surveys was to analyze the quality of the data produced on different attributes of persons, dwellings and buildings. However the problem of under- and overcount was still relevant for dwellings and buildings because of the shortcomings of these data in the CPR.

The quality surveys of the 1985 census fall into two categories: those analyzing the quality of the CPR data (especially the data on household-dwelling units and dwellings) and those analyzing the census form data on the economic activity of the population. The data on household-dwellings units and dwellings, for instance, were analyzed by comparing the census data with corresponding data from the 1985 Household Budget Survey,

an interview-based sample survey of 12,000 households. Another source of dwelling data was a sample survey of dwellings registered as unoccupied in the CPR. This survey provided information on the overcount of dwellings in the CPR.

The quality of the data on the economic activity of the population was analyzed with the help of processing error studies (data entry errors and errors in coding and editing) and a special quality survey in which the final census data on persons were compared with the checked and corrected data of the interview-based regular Labour Force Survey. Some experiences from this survey and the methodology of the survey will be discussed in this paper. A short general description of the regular Labour Force Survey will be presented in the next chapter.

The Finnish monthly Labour Force Survey

The Finnish Labour Force Survey (LFS) is a sample survey based on a random sample of 12,000 persons selected from among the population aged 15-74 years. Data collection takes place mainly in personal interviews carried out by the CSO's interview organization.

The person interviewed is asked questions about his labour force participation (current activity), employment, unemployment, workplace, occupation, industrial status, time use, days and hours actually worked, overtime and secondary jobs, and normal hours of work. About 94 per cent of the interviews are telephone interviews and five per cent personal interviews. About one per cent of the answers are obtained using a mail questionnaire. The average non-response rate of the survey is about 4.7 per cent.

Structurally, the survey is a so-called continuous panel survey. The monthly sample of 12,000 persons is broken down into five rotation groups, each of which contains 2,400 persons. In the monthly survey, each rotation group can be considered to be an independent random sub-sample of the whole monthly sample. Each rotation group is surveyed five times over a period of 15 months. The lag between interviews is three months, with the exception of the lag between the third and the fourth interview which is six months. For estimation purposes, the sample is post-stratified by geographical area, age group and sex. The reference period of the survey is one week.

The Labour Force Survey as a frame of the quality survey

The monthly LFS made an excellent frame for the quality survey of the census form data for the following reasons:

1. As pointed out above, the census and the LFS used the same concepts and definitions in measuring the current activity of the population.
2. The study week of the LFS coincided with the census week.
3. The high response rate of the LFS, combined with more thorough interview questions and manual process-

ing by a more highly qualified personnel, made it reasonable to assume that, on the average, the final LFS data on persons were closer to the "true" values than the corresponding census data.

4. The uniform personal identifier used in Finland allowed easy and errorless linkage of the census records with the corresponding LFS records.

One important reason for using the LFS was that it was cheaper than a special separate quality survey like the one carried out in connection with the 1980 census.

The principal variables of the census selected as the topics of the quality survey were as follows:

1. Current activity of the population

subdivided into

Labour force

Employed

Unemployed

Persons not in the labour force

Conscript members of the defence forces

Students and pupils

Pensioners

Home-makers

Others

2. Industrial status

subdivided into

Wage and salary earners

Own-account workers

Employers

Unpaid family workers

3. Employer sector

subdivided into

Private sector

Central government (incl. provincial government)

Local government)

4. Industry

A national version of the ISIC.

5. Occupation

A national version of the ISCO.

Because of the quality survey, it was necessary to take a

few extra measures to check and to better the LFS data in order to make the "true" values as good as possible. In order to further reduce the cost of the quality survey, these extra measures were restricted to the first LFS rotation group, i.e., the sub-sample of 2,344 persons not interviewed earlier in the LFS.

The extra measures were as follows:

1. Additional questions (a total of 70, in addition to the standard 25 LFS questions), specially formulated from the viewpoint of the quality survey, were asked by the interviewers in connection with the LFS data collection. The collection of the data of the first rotation group was delayed by a few days in order to allow the group to respond to the census questionnaire before the LFS interview. The non-response rate for the additional questions was 10.2 per cent, compared with 5.4 per cent for the regular LFS questions asked of the first rotation group.

2. After the usual data processing operations, the data of the first rotation group were rechecked and recoded by persons not involved in the original checking and coding of the questionnaires of the first rotation group. The data from the additional questions were utilized in this process.

The results of the rechecking and recoding demonstrated the overall high quality of the LFS data. Concerning current activity, only 0.7 per cent of the original LFS values had to be corrected. The situation was similar with respect to industrial status. When it comes to the industrial and occupational classifications, the percentage of erroneous codes was somewhat higher, 4.3 and 4.0 per cent, respectively, of the number of employed persons.

Estimation of "error" in the census data

The usual procedure in this kind of quality survey setup is to match the quality survey records with the census records of the corresponding persons and to calculate the gross and net errors of the study variables as follows:

Gross error of a variable =

100 per cent minus the percentage of similarity classified cases

Net error of a variable =

the relative difference between the marginal distributions of the variable in the census and those in the quality survey.

This procedure could be applied directly to the sample of the first LFS rotation group. In this case, however, there were even uncorrected data on the whole LFS sample which correlated highly with the corrected "true" values of the first rotation group. Therefore, error estimates could be compiled using the so-called difference estimator (a special type of linear regression estimator see e.g. Madow, W.G.: "On some aspects of response error measurement", Proc. of Social Stat. Sc., American Statistical Association, 1965, pp. 182-192).

The difference estimator is of the form

$$Y = y + (x' - x)$$

where

x = the percentage of cases where the census values are identical with the original values of the first rotation group

y = the percentage of cases where the census values are identical with the checked and corrected values of the first rotation group

x' = the percentage of cases where the census values are identical with the original values of the whole LFS sample

The estimate of the gross error of a variable is given by $1 - Y$ and the variance of the estimator is of the form

$$V(Y) = \frac{y}{1/n - 1/N} S_y^2 + (1/n - 1/n') (S_x^2 - 2S_{xy})$$

where

n = the sample size of the first rotation group

n' = the sample size of the whole LFS

S_x^2 = the population variance of variable x

S_y^2 = the population variance of variable y

S_{xy} = the population covariance of variables x and y .

In this way, the average of the standard errors of estimates could be reduced by about 50 per cent as compared with the standard errors for the sample of the first rotation group only. To have reached the same level of standard error by ordinary estimation methods, at least four times as large a sample should have been asked the additional interview questions and, consequently, at least four times as many data should have been rechecked and recoded.

¹Presented at the IFDO/IASSIST 89 Conference held in Jerusalem, Israel, May 15-18, 1989.

Post Censal Surveys in Great Britain

by Robert Barnes¹

Office of Population Censuses and Surveys
London

Introduction

Censuses of population have been carried out in Great Britain every ten years since 1801 with the exception of 1941. In addition a mid-term census was conducted in 1966 on a sample of 10% of the population. The next census is being planned for April 1991. Censuses in Britain are still carried out in a conventional way in that data are collected through enumeration procedures designed for the purpose rather than from registers. Specially recruited enumerators are used both for the delivery and the collection of forms.

Post censal surveys are of much more recent origin - the first having followed the 1961 Census. They have been of three kinds:

- post-enumeration surveys to evaluate the coverage and quality of the census;
- follow-up surveys which use the census as a frame from which to draw samples of groups of the population for more detailed enquiry;
- a longitudinal study which for a sample of about 1% of the population links data from successive censuses together with vital events registered during the inter-censal period.

Post enumeration surveys (PES)

The first post enumeration survey was carried out in 1961. Its aim was to assess the coverage and quality of the 1961 Census in England and Wales. Although separate from the main census, the checks were carried out by census personnel. The sample for the coverage check comprised a systematic random selection of 2,500 enumeration districts (EDs) as first stage units. Within each a 'plot' was identified containing roughly twenty addresses bounded by features that could be identified on the ground. A shortcoming of the system was that it did not provide adequate checks of coverage error arising at the boundaries of enumeration districts. The check produced an estimate of net under enumeration of 0.2% but the General Report on the 1961 Census² said "The design of the enquiry was such that the quality of the result may be suspect but there is no information on this".

The quality check sample was selected from the same

'plots' as the coverage check. In all about 17,500 addresses were revisited for voluntary interviews within about three weeks of the census date. Census enumerators were used for this purpose but were not skilled interviewers. Again the General Report comments on the shortcomings of the study. "This fact and the very limited instruction which it was practicable to give them were contributory factors in the failure of the post-enumeration survey to give satisfactory answers to some of the questions that were included".

In 1966 a quality check on the sample census was carried out by the Social Survey Division of the Central Office of Information. This was a government organisation separate from that responsible for the census itself. A sample of just over 5,200 households was selected from 300 EDs in England and Wales. There were several key features of this quality check survey which were to set the pattern for similar studies carried out in conjunction with future censuses.

- Whereas cooperation in the census was, and still is, compulsory, the post-enumeration survey was voluntary. All sample surveys of households and individuals in Britain are voluntary. Response to the 1966 Quality Check was 95%.

- Whereas the census employed a large force of temporarily recruited enumerators, the survey used highly trained and experienced interviewers.

- The survey used detailed questionnaires to ascertain the 'true' answer to topics which in the census were covered by just one or two questions.

- Survey interviewers had copies of the informant's answers on the census form so that they could probe any discrepancies. This not only improves the quality of the check but also provided some reasons for the differences.

- The survey sought interviews from each adult in a household and so relied far less on proxy responses than did the census.

- In order to avoid discrepancies arising through genuine change between the time of the census and the time of the survey, it was important that the survey took place soon after the census. In fact the survey fieldwork was carried out between two and three months after the

date of the census.

The results of the quality check were published³ and indicated that answers to a number of questions on the census form were in error to a substantial degree. The worst item was the question on the number of rooms in a household - this was misclassified in approximately one case in five.

Coverage and quality checks were again carried out on the 1971 Census. As before the coverage check was conducted in England and Wales by census officers. Although the check showed an undercount of only 0.23%, the check was not considered entirely successful from a technical point of view. The General Report on the 1971 Census⁴ said "it was carried out by census officers some of whom would still have been busy with other work on the census. Some census officers might have viewed the check as a fault finding mission and in consequence would not have been fully motivated to ensure its success".

In fact it was demonstrated that the accuracy of the coverage check was poorer than that of the census itself since there were more addresses found in the census but missed by the coverage check than vice versa. Neither did the coverage check provide a reliable measure of the extent of over counting - that was provided by other means. Taking account of the deficiencies of the check it was finally concluded that a more realistic figure for the net undercount was about twice that shown by the check.

The 1971 quality check was conducted by means of a post enumeration survey, covering the whole of Great Britain, by interviews at just under 5,000 addresses. The response rate was 85%. The check was again carried out by the Social Survey which had in 1970 become a division of the same Office which was responsible for the census - the Office of Population Census and Surveys. As with the coverage check there were unsatisfactory aspects about the quality check. It was originally planned that fieldwork should be completed within about two months of the census. In the event the work did not start till then and, because of difficulties in transferring census information from one division to the other, it proceeded slowly. Eventually fieldwork was completed about five months after census night. Because of weaknesses in the design of the check and the length of time to complete fieldwork, a number of important aims of the check were not achieved. For example there was no assessment possible of the accuracy with which questions on household tenure and amenities were answered.

In 1981 coverage and quality checks were carried out but this time both were undertaken by the Social Survey Division of OPCS. This provided a much more integrated approach to census evaluation.

For this coverage check roughly 1,000 EDs (about 1%) were selected in England and Wales⁵ and all addresses in them were thoroughly relisted by trained interviewers to see if any had been missed by the census enumerators. The EDs were selected in blocks of four, adjacent to each other so that enumeration of ED boundaries could be

checked. Different samples of addresses were selected within the EDs to see whether persons had been left off census forms for addresses otherwise correctly enumerated, to see whether anyone had been missed in addresses enumerated as vacant or non-residential and to check especially the enumeration of multi-household addresses.

This was a much more rigorous approach than had been used in 1971. Moreover in 1971 no attempt had been made in the field to reconcile discrepancies. In 1981 the interviewers were given copies of census listings and so could check by means of personal interview numbers of persons missed in non-enumerated addresses. Also in 1981, unlike previous checks, discrepancies in the number of persons present in enumerated households were taken up with the informant, enabling the interviewer to form a more definite idea as to whether the census form or the post enumeration information was correct. There was one other way in which the 1981 check was superior to that of its predecessors. All EDs were graded on the basis of 1971 census data for expected difficulty of enumeration. The design used for the 1981 post enumeration survey over sampled, by a factor or two, these 'difficult' EDs, since it was hypothesized that such EDs would be likely to produce more errors.

However in spite of the thoroughness with which the 1981 coverage check was carried out and the improvements made over previous such studies, there are inherent difficulties in checking the coverage of a census using re-enumeration methods. Because of the cost of the approach, the samples have to be fairly small and the sampling errors are therefore relatively large - especially at the sub-national level. Also in spite of the thoroughness of the methods it is likely that some of the persons and addresses missed in the census will have been missed in the survey too - for similar reasons. Therefore in 1981, as in previous census evaluations, in addition to the post enumeration survey, checks were made also against independent administrative sources for particular groups of the population. For example checks were made for children aged 0-9 with data on registered births and deaths and making allowances for migration, for infants aged 0-1 with birth records, for children of school age with numbers on school rolls, and the census count of people of retirement age was compared with the records of the numbers receiving State Retirement Pensions. These checks confirmed that the level of under enumeration in the 1981 Census was small even though it may have been slightly higher than the half per cent found by the post enumeration coverage check.

In addition to the post enumeration check for coverage, a sample of about 5,000 households throughout Britain within the EDs selected for the coverage check was revisited for detailed interviews to check the accuracy of answers given on census forms. Response to this post enumeration quality check survey was over 90%. All fieldwork was completed within three months of the date of the census. The 1981 census form was the shortest for fifty years. Compared with many items which might have been included, and which are in other countries, the items in the British census might have been regarded as

relatively straightforward and commonplace. Nevertheless gross error rates of 8% or more were found on five of the sixteen variables examined. For two of these the error was over 25% (and again the worst case was the question on number of rooms) and for some sub-groups the gross errors were even larger. The relative importance of these errors depends on the uses made of the information - for example the error rate reduced substantially when classifications were collapsed into fewer and simpler categories and, in any case, the net errors, in distributions, were smaller.

Improvements were made for the 1981 post enumeration survey in regard to publication of the results, compared with previous censuses. The full report⁶ took a few years to produce but key results were published more quickly in the form of summary monitors both for the coverage check and the quality check⁷.

The 1981 Census quality check also provided an opportunity to carry out a study to measure the coverage and quality of the electoral register.⁸ This had first been done, although in a more limited way, in conjunction with the 1966 quality check. The 1981 study covered the whole of Great Britain and measured the extent of both persons who were eligible to be on the electoral register but who were not, and those who were on the register but for whom there was no census form or for whom census details suggested they were not eligible.

The results showed that just under 7 per cent of those who should have been on the register were not and that the same proportion of names on the register should not have been. The under representation was especially high in Inner London (14 per cent), and among those aged 17 (24 per cent), and those who had recently moved (27 per cent). These figures represented a deterioration since the register had previously been checked in 1966. It is planned to carry out a similar check in 1991.

Follow-up surveys

The first survey to use the census as a sample frame for more detailed enquiry was a follow-up to the 1966 sample census. This was a study on diet and health concentrating especially on sugar intake because of the supposed relationship between that and myocardial infarction. This was a postal enquiry of some 20,000 persons conducted in 1967/68. 1966 sample census returns were used as a sample frame because the sample required was of men between ages 45 and 65 in the London area. Moreover the subsequent fate of the sample members could be traced through death registers also maintained by the same Office. Considering the length of time between the census and the survey, the percentage of returns was high at 85%. Of these 89% had completed the forms (75% response overall). Of the remainder the majority were returned as gone away, deceased etc. Important aspects of this survey were that data were collected by the same organisation that carried out the census and that procedures to ensure the anonymity of sample members were strictly adhered to before the data records were released to outside researchers for analysis.

The tracing of sample members through death registers has continued ever since and, although no results have yet been published on the relationship between sugar and heart disease the study has yielded other results relating diet and disease - in particular between tea and coffee consumption and cancer.⁹

Following the 1971 Census there were three such enquiries. The first was the Income Follow-up Survey.¹⁰ There had been pressure to include an income question in the census itself but tests carried out in 1968 and 1969 indicated the severe problems of dealing with such a complex topic with simple questions suitable for a census form. It was also found that the inclusion of income as a possible topic in a compulsory census aroused hostility among some sections of the British public. Therefore it was decided that the Census division of OPCS should carry out a voluntary survey, by post, on 1% of the population throughout Great Britain in such a way that the answers could be linked with the census forms of the sampled individuals. Although it was intended to carry out the study as quickly as possible after the census, the adverse publicity which the 1971 Census attracted caused the follow-up study to be postponed and it was eventually conducted over a year later. Because of this and because of the subject of the enquiry, the response was only 40% and a substantial volume of imputation, using hot deck methods, was carried out. The study was not repeated after the 1981 Census.

A second follow-up survey to the 1971 Census was on the subject of qualified manpower. This also was a voluntary, postal enquiry carried out by Census division. The object was to follow up a sample of those who on the census form had reported that they had academic, professional or vocational qualifications, to seek more information about the qualifications, their jobs and their employment income. However because of competing demands of other work, there were delays in completing this survey. In the end the need for the results was overtaken by other events and the report was not published.

The third follow-up to the 1971 Census was the so-called Nursing Survey.¹¹ This was particularly important not only because of the topic of enquiry but because the criticism which the survey attracted had important consequences for follow-up enquiries of future censuses. The purpose of the study was to obtain information from persons with nursing or midwifery qualifications, who were of working age but who were not practising nursing or midwifery. The census provided an ideal frame from which to select a sample of such persons who would otherwise would have been especially difficult to locate. A sample just over 700 throughout Great Britain was selected for voluntary face-to-face interview and of these 89% agreed to cooperate. The fieldwork was carried out within four to five months of the census date.

The study was carried out by the Social Survey Division of the then recently formed OPCS. Census Division, of the same Office, selected the sample and passed the names and addresses to the Social Survey. Because both

divisions were part of the same Office there was no breach of the confidentiality undertakings given at the time of the Census. But this fact was not always appreciated by the public, nor by some sections of the press or even the research communities and there was considerable criticism of the practice of using census forms for this kind of purpose. Even an investigatory team from the Royal Statistical Society commented in 1973 that "This use was, in our opinion, only doubtfully covered by the wording printed on the census form and by other public pronouncements on the confidentiality of information given in the census operation". Nor did the reference to the nursing survey was still being made at the time of, and after, the 1981 Census even though there were no follow-up enquiries of this kind attached to that Census.

As a result of the controversy, procedures changed after 1971. Although there were no such studies following the 1981 Census, had there been they would have had to have been announced to Parliament before the census was carried out. Similarly for 1991, the White Paper published in July 1988¹² announcing the government's intentions to take a census 1991, contained the following paragraph.

"The census may be used as a source from which to select samples for further more detailed surveys, for example of people with particular educational qualifications. Response to any such survey would be voluntary.... Information would be treated in the same strict confidence as information given on the main census forms. It is too early to know whether there will be a need for any such census-linked surveys and the topics that might be covered, but Parliament will be informed before the census is taken about the subject matter of any census-linked survey which it was proposed to conduct following the census and all those completing the census forms would also be informed of the possibility of being asked to participate in such a voluntary survey after the Census."

The White Paper also emphasized that any such census-linked surveys would have to be handled by OPCS (and by the General Register Office in Scotland). This means that the work cannot be contracted to commercial or academic research agencies.

The longitudinal study¹³

This is a rather special form of post-census survey. It is a data linkage exercise and involves no specific enquiry of the sample members. The study comprises a sample of persons in England and Wales with birthdays falling on each of four specific days in the year and therefore provides roughly a 1% sample of the population. The study started with the 1971 Census records for the sample members. Between 1971 and 1981 mortality and migration data were added to the records from registration data also held by OPCS. Also added to the records were live and still births to sample members, deaths of children under one year of age to sample members, death of a spouse and cancer registration. In addition children born on any of the four sample days during any year

between the censuses were added to the sample, thus maintaining it at approximately 1%. In 1981 the new intercensal data on migration and vital events has continued to be added since then. Proposals are now being drawn up for further links to be made to the 1991 Census.

Initially the longitudinal survey included about 530,000 people selected from the 1971 Census. By the time of the 1981 Census some 121,000 more had been added to the sample because of new births, immigration, or because they were found for the first time in 1981. And 115,000 were removed from the sample because of death, emigration or because they could not be traced in the 1981 Census. Therefore the sample after the 1981 Census records are available, numbers in particular sub-groups are as follows:

<u>Characteristics in 1971</u>	<u>Number for whom 1981 details linked</u>
Children under 16	120,000
Teenagers 13-19	40,000
Divorced	5,000
Lone parents with dependent children	5,000
Men aged 50-64	40,000
Persons aged 75 or over	10,000
Unemployed	10,000
Migrants	130,000

There are many important uses for this longitudinal data set for such a large sample of the population. They are too numerous to list here but they include the relationship between socio-economic circumstances and mortality in relation to housing and household circumstances, socio-demographic differentials in cancer incidents and survival, the relationships between socio-demographic factors and migration (i.e. the propensity to change address about the time of events such as marriage, birth of children, birth or bereavement), fertility patterns of young women according to family background characteristics, and social and occupational mobility. Apart from this kind of general analysis, the data set is also large enough to focus on particular sub-groups or geographical areas.

Conclusions

Post census surveys have been carried out in Great Britain for nearly 30 years. They have been used to evaluate the census,¹⁴ to take advantage of the census as a sampling frame and to provide longitudinal data for particular topics. They have not all been successful and problems both technical and ethical have arisen. However post census surveys will almost certainly be a feature

of future censuses in Britain and lessons learned from previous experiences should help to ensure that past difficulties are avoided. The main lessons are that field work for follow-up studies should be completed as soon as possible after the census, that the field workers should be trained interviewers skilled in administering complex questionnaires, that the public and Parliament should be informed in advance of the intention to use census data for this kind of purpose, and that census evaluation studies should be carried out by staff who were not directly involved with the census operation itself. □

¹Presented at the IFDO/IASSIST 89 Conference held in Jerusalem, Israel, May 15-18, 1989.

²General Report on the 1961 Census.

³Gray & Gee. A Quality Check on the 1966 Ten Per Cent Sample Census of England and Wales. HMSO, 1972.

⁴Census 1971. General Report Part 3, Statistical Assessment. OPCS, 1983.

⁵Separate coverage checks were carried out in Scotland.

⁶Britton & Birch. 1981 Census Post Enumeration Survey. HMSO, 1985.

⁷Evaluation of the 1981 Census. OPCS Monitor CEN 82/3, 1982.

Evaluation of the 1981 Census: Post Enumeration Survey. OPCS Monitor CEN 83/4, 1983.

Evaluation of the 1981 Census: Post Enumeration Survey (Quality Check). OPCS Monitor CEN 84/3, 1984.

⁸Todd & Butcher. Electoral Registration in 1981. OPCS, 1982.

⁹Leo Kinlev et al. Coffee and pancreas cancer; controversy is part explained. *The Lancet*, Feb 1984.

L.J. Kinlev et al. Tea consumption and cancer. *British Journal of Cancer*, no. 58. Macmillan, 1988.

¹⁰1971 Census Income Follow-up Survey. OPCS Studies on Medical and Population Subjects, no. 38. HMSO, 1978.

¹¹Sadler & Whitworth, Reserves of Nurses. HMSO, 1975.

¹²1991 Census of Population. Cm 430. HMSO, 1988.

¹³The Longitudinal Study 1971-1981. Cen 81Ls. HMSO, 1988.

Brown & Fox, OPCS longitudinal study: ten years on. *Population Trends*, no. 37. HMSO, Autumn 1984.

¹⁴Whitehead. The GRO use of social surveys. *Population Trends*, no. 48. HMSO, Summer 1987.

The Canadian Experience with Post Censal Surveys

by Adele D. Furrie, Program Manager
Post-Censal Surveys Program
and Craig McKie¹ Editor
Canadian Social Trends, Statistics Canada

Abstract

The Canadian experience with post-censal survey is, at this point in time, limited to two surveys - one which used the 1971 Census returns as the sampling frame, and the 1986 survey, which used the 1986 returns as the sampling frame, utilized the Census field organization to collect the data, and used the 1986 Census data to supplement the data collected in the post-censal survey. Post-censal surveys provide efficiencies in terms of overall costs because of the accessibility of the Census data to identify relatively rare populations. The availability of a field organization to a) select the sample immediately following the collection of the Census data and b) collect the data reduces the overall cost to the post-censal program since the hiring and some of the training costs are absorbed by the Census. The availability of the Census data, not only for the population of interest, but for the data base. Respondent burden is reduced because many of the demographic and socio-economic variables are included in the Census. This is true also for family and household-related data.

Some consideration is being given to the conduct of a similar post-censal disability survey following the 1991 Census. In addition, there is the possibility for a survey on the senior population and one on aboriginal persons.

The Canadian experience with post-censal surveys is limited to two surveys - the Highly Qualified Manpower Survey (HQMS) conducted following the 1986 Census. While the methodology of the two surveys differed significantly, each provides an interesting application of post-censal methodology which could be considered by those who are planning such activities. The methodology for each survey will be described and an evaluation will be provided of the methodology.

The Highly Qualified Manpower Survey - HQMS - was conducted in the fall of 1973, just over two years after the 1971 Census was conducted. The objective of the survey was to assess past expenditures for education in relation to the utilized labour force status. The data were needed to assist in the formulation of policy relating to long-range planning in the fields of education and manpower planning.

The sample was selected from the Census database based on the individual's age-sex-labour force status and level

of post-secondary attainment. Because name and address were not part of the Census database, the original Census questionnaires had to be accessed, and information such as telephone number and the name of the head of the household as well as the name and address of the selected person was transcribed. Quality control procedures for the transcription and creation of the name and address file were employed to control for errors.

The questionnaire covered a limited range of topics such as field of study, current labour force status and current earnings, and an employment profile over time. These data were supplemented by the information collected in the 1971 Census, so that the combined database covered a wide range of topics, such as ethnicity, immigration status, marital status, etc.

The survey questionnaire was mailed to approximately 138,000 selected persons (the number of persons meeting the selection criteria were approximately 720,000); about 72,000 or almost 70% returned the completed questionnaire. A tracing operation was conducted to establish the current address for those selected persons who had moved. Follow-up of non-respondents was conducted by mail, and in some instances, by personal visit, during the period September, 1973 to March, 1974.

Because the survey was conducted over two years after the Census, there was some difficulty in locating some of the selected persons and some of the data retrieved from the Census were out-of-date. For example, the individual may have married or additional children may have been born, and these differences would not have been accounted for in the combined database.

The 1986 post-censal survey, the Health and Activity Limitation Survey (referred to in this paper as HALS) employed a different collection methodology, taking into account the experience of the HQMS. HALS was comprised of three distinct surveys - two household surveys - both of which used the disability question on the Census as the screen to identify the sample and an institutions survey which used the Census to identify the location, size and type of the institution.

The need for a comprehensive database on disabled persons was articulated in 1981 in the report entitled *Obstacles*, the report from the Special Parliamentary Committee on the Disabled and the Handicapped. It

noted that there were no national data available on disabled persons. Not only were there no national estimates on the number of disabled persons in Canada and the nature and severity of their disability, but, as important, there were no estimates on the barriers which these disabled Canadians face in the conduct of their everyday activities. Because many of the programs and services offered to disabled persons are the responsibility of provincial and local governments, it was important that these data be available for relatively small geographic areas. As programs and services differ for different age groups, the sample would also have to be large enough to be able to generate estimates within the geographic areas for each age group. The methodology of a post-censal survey was considered as the only viable alternative to obtain this level of detail. Other options such as the use of existing survey vehicles - the monthly Labour Force Survey or the annual General Social Survey - were considered but it was determined that neither would yield a sufficiently large sample of disabled persons. There were also limitations in the coverage for both of these survey vehicles. The Labour Force Survey excludes the more remote areas of Canada, residents of Indian reserves, and residents of institutions. The General Social Survey generally utilizes a random-digit dialing methodology to create a sample, therefore, households without telephones would be excluded from the survey. The General Social Survey is also household based so that residents of institutions would be excluded from the survey.

The decision to use the Census as the method to identify the samples for the household surveys necessitated the inclusion of a disability question on the Census questionnaire. It was decided that this question would be added to the "long" questionnaire - the one completed by one out of every five households. The question asked if the individual was limited in the kind or amount of activity he/she could undertake because of a health problem or condition. The second part of the question asked if the individual had any long-term disabilities or handicaps. Households were advised through the Guide that was included with their Census questionnaire that the disability question was to be used to identify a population for a more-intensive survey on the issues facing disabled persons.

It was determined through a pre-test that this question would identify most of the more-severely disabled population, but that additional questions would be required to identify all disabled persons. A copy of the Census disability question is included in Appendix A of this paper.

The content of the HALS questionnaire was determined through extensive consultation with representatives from government departments that provided programs for and services to disabled persons. Consultation with organizations of and for disabled persons was also undertaken to ensure that their needs were reflected in the content. The questions used to identify the nature and severity of the individual's disability were, for the most part, developed by the O.E.C.D. These questions are known as the

Activities of Daily Living and were developed to identify physical and sensory disabilities. Other questions were added to identify emotional, psychological and learning disabilities and persons who are developmentally delayed.

With the inclusion of the disability question on the Census, the sampling frame was in place for the post-censal survey of disabled persons. To maximize the efficiency of this sampling frame, it was decided that an operation should be integrated with Census processing to select a sample of individuals who had responded "Yes" to the Census disability question. This would allow for the conduct of the post-censal survey shortly after the Census, thus minimizing the follow-up required because of inter residence moves taking place between the time of the Census and the conduct of the post-censal survey. It would also enable the utilization of the Census field staff to conduct the face-to-face interviews.

To accommodate the selection of the sample and to ensure that the field staff would be available for further work, geographic areas were identified prior to the conduct of the Census. These areas were defined as the geographic area within which the workload for one Census Field staff (Census Representative) was located. Census staff received additional training on the concepts and definitions used in HALS and the face-to-face interviews were conducted immediately following the Census collection. The reference day for the 1986 Census was June 1. In most instances, the interviews were completed during August and September, 1986. There were an estimated 120,000 individuals selected for the follow-up interview; the overall response rate was in excess of 95%. The second household sample involved a sample of individuals who had responded "No" to the Census disability question. This sample was necessary because the pre-test had indicated that some disabled persons may not respond affirmatively to the general disability question. A sample of approximately 80,000 individuals was selected during a later stage in the Census processing, but before the Census documents were returned to Head Office in Ottawa. The same questionnaire was used and most respondents were contacted by telephone. Those who did not provide telephone numbers on their Census questionnaire were contacted in person. The survey was conducted from the Regional Offices of Statistics Canada during October and November, 1986 by interviewers who are part of the regular Regional Office staff. Approximately 90% of the sample was contacted and agreed to participate in the survey.

The data from both household surveys was integrated with the Census at the micro-record level, so that the linked database contains information from both the Census and HALS. The Census data, for the most part, is for the selected individual, but included in the base are also some variables about the family and household within which the selected person resides. Because the Census and HALS were conducted within six months of each other, the variables taken from the Census such as marital status should not have changed significantly.

Another feature which adds to the richness of this database results from the sample being selected to represent both "Yes" and "No" respondents to the Census disability question. The HALS sample can be divided into strata - those who are disabled and those who are not. For the non-disabled population, the data available on the linked database includes all of the Census variables. For the disabled population, the data includes both Census and HALS variables. This affords the user the opportunity to make comparisons of the characteristics of the disabled and the non-disabled populations.

The Census methodology did not include the use of the long questionnaire in institutions; therefore, the disability question was not asked in institutions. To obtain information for residents of institutions, the Census was used to identify the location, size and type of institution. Penal institutions and correctional facilities were excluded because of operational difficulties. From the remaining list, a sample of institutions was selected, approximately 1,100 out of a total of approximately 5,300. Each of the selected institutions provided a list of residents from which a sample was selected. Of the 18,200 residents selected from the 1,100 institutions, less than 3% refused to participate in the survey. A personal interview was conducted with slightly over 50% of the sample. For the remaining sample of respondents whom the institution administrator deemed to be too ill or too disabled, the interview was conducted with an individual who provided the day-to-day care.

Data from HALS was released in May, 1988 and has been used by both the public and private sectors. Much has been learned concerning the conduct of post-censal surveys and the integration required with the Census operations.

Planning is now underway for post-censal survey activity following the 1991 Census. Based on consultation with representatives involved in social programs, three potential topics have emerged and further consultation is now underway. The three topic areas are a survey of seniors with the focus on support networks, a repeat of the survey of disabled persons so that data are available over time, and a survey of aboriginal persons, both on- and off-reserves. The possibility of one or more of these topics going forward is contingent on obtaining funding for them. The 1986 survey cost seven million dollars. That survey proved that a post-censal survey, closely linked to the Census operation in terms of identifying the sample, utilizing the Census field staff, and the Census data is a viable option for surveys of relatively rare populations or which require significant geographic detail. □

Appendix A

20. (a) Are you limited in the kind or amount of activity that you can do because of a long-term physical condition, mental condition or health problem:

At home?

No, I am not limited

Yes, I am limited

At school or at work?

No, I am not limited

Yes, I am limited

Not applicable

In other activities, e.g. transportation to or from work, leisure time activities?

No, I am not limited

Yes, I am limited

- (b) Do you have any long-term disabilities or handicaps?

No

Yes

¹ Presented at the IFDO/IASSIST 89 Conference held in Jerusalem, Israel, May 15-18, 1989.

Major Post Censal Redesign of Household Sample Surveys in the United States

by Preston Jay Waite¹
U.S. Bureau of the Census

Acknowledgements

I wish to acknowledge the work of many staff members in the Statistical Methods Division of the U.S. Bureau of the Census in the preparation of materials for this paper. Special thanks to Dr. Charles Alexander and Gary Shapiro for their contributions to the editing and review and to Pat Curran for her work in preparing the manuscript.

Post Census Surveys (PCS) are utilized in a variety of ways in the United States. A large sample (one household in six) is imbedded in the Census collection itself. This sample allows for the collection of detailed housing and persons information not covered by a complete count. We also use sampling for the measurement of undercoverage by selecting a large post-enumeration survey (PES) of approximately 150,000 households. The results of this survey are then matched to the census enumeration to measure the extent and characteristics of census undercoverage. Major census follow on surveys of Residential Finance and of scientists and engineers are conducted immediately following the census. Frames for these surveys are constructed by screening units with particular characteristics from census questionnaires.

All of these survey collections are major operations and complete papers suitable for this conference could have been produced for each of them. I would like to focus my remarks today, however, on an additional use of the census that being for a frame for selection of the major household surveys conducted by the United States Government.

This paper will discuss our ongoing plans to redesign our current household surveys based on the 1990 census. I will discuss our general methodology and the challenges we face by trying to simultaneously select several surveys simultaneously. I will also mention briefly some of the planned uses of new technologies in the reselection of our survey samples.

Using the Census As A Frame For Continuing Household Survey

The United States Decennial Census Address Lists are used as a sampling frame for many of the Government's major continuing household surveys. The principal household surveys using the census as a frame are:

1. The Current Population Survey, (sponsored jointly by the Labor and Commerce Departments; the basic labor force survey).
2. The Consumer Expenditure Surveys, (sponsored by the Labor Department; used as input to the Consumer Price Index).
3. The Current Point of Purchase Survey, (sponsored by the Labor Department; consumers are interviewed to generate a frame of retail outlets for measuring prices for the Consumer Price Index).
4. The Survey of Income and Program Participation, (sponsored by the Commerce Department; a longitudinal survey which follows persons every four months for two-and-one-half years to collect information on income dynamics and use of government transfer payments programs).
5. The National Crime Survey (sponsored by the Justice Department; collects information from victims of crime).
6. The American Housing Survey (sponsored by the Department of Housing and Urban Development; a biennial longitudinal survey of housing which updates the census data for sample units, while adding in new construction).

The census address lists are the primary source of the sample for all of these surveys. These lists give us a frame for the United States as of census time 1990. Since these are continuing surveys, the sampling frame derived from the census must be kept up to date between censuses. This is done mainly by sampling building permits, which are required for new construction in most parts of the country. Permits are listed and sampled every month from selected building permit offices. Where such permits are not required by local governments, new construction is represented through area sampling. In the area sample, a list of all the addresses for selected areas on the map is made by the field representative. Area sampling is also used for both old (existing in 1990) and new construction in some mostly rural areas where permits are not required for new construction and for areas where the census addresses are hard to locate.

The census address list is an inexpensive source of sample, compared to an area sampling approach, and it

gives more complete coverage of the population than any other available list of addresses. But even with the census list we find coverage to be a problem. Potential coverage problems can be of two types; coverage of households and coverage of persons within households.

Evaluation of the coverage of households in the census shows that at most a one to two percent overall under-coverage at the time of the census, although undercoverage for households of minority races is known to be substantially worse than for the population as a whole. Coverage of households by the continually updated census frame is more difficult to measure, but estimates of missed households range from about one to five percent. Estimates of missed persons are more reliable, since the survey estimates of the number of persons may be compared to updated census estimates. This comparison shows on the order of a 10 percent undercoverage of persons, with the worst coverage for young males. There is evidence that for young males most of this is due to failure to obtain complete lists of household members, rather than to missing households.

The updated census estimates of persons by age, race, and sex are produced by inflating or deflating the census counts for births, deaths, immigration and emigration. Most of the survey estimates are calculated using post-stratification to bring the final survey estimates of persons into agreement with the updated census estimates.

The surveys using the census as a frame are all conducted by the Bureau of the Census, although the data may be analyzed and published by other government agencies or research organizations. By law, no one outside the Census Bureau may have access to the actual census addresses, nor to information which would permit the identification of any sample household. This places limits on the amount of detail which can be included on data files intended for public use. It also means that only sworn Census Bureau agents may contact the sample households for any survey which uses the census as a frame. To avoid these restrictions, another major household survey conducted by the Census Bureau, the National Health Interview Survey, uses only an area sample. The operations for this survey are coordinated to some extent with the other household surveys, and it is redesigned in conjunction with them, but the frame for this survey is created independent of the census.

Having a single field staff conduct the interviews for all survey is extremely cost-effective. The administrative and office costs can be shared. Also, in many cases, the same field representative can conduct interviews for several surveys, since the surveys take place at different times of the month. This sharing of field representatives reduces the number of field representatives who have to be recruited and trained. Detailed coordination of the sampling operations also saves effort and money. The lists of building permits used to keep the frame up to date can be shared among the surveys. In the area frame, the later surveys in a particular map area can make use of the lists made for the earlier surveys. The cost savings from sharing listings are substantial. Since the sample ad-

resses for the different surveys are kept close together whenever possible.

Redesigning the Sample After Each Census

The census address list was first used as a sampling frame for the ongoing household surveys following the 1960 census. In theory, this frame could have been kept up to date perpetually by adding new construction from building permits and area listings. In reality, the sample was reselected after the 1970 census and again after the 1980 census. The sample will be reselected after the 1990 census. One reason for reselection is the likelihood that in spite of our best efforts, continued updating of the old frame inevitably will lead to a gradual deterioration of coverage. Additionally, as time goes on, a greater proportion of the sample would come from the more expensive permit frame.

A basic reason for redesigning the sample after each census is to use information from the new census in improving the design. The census information collected for each household includes household size, race of the occupants, whether the unit is a farm, whether the dwelling is rented or owned, and the rent or value of the dwelling. A sample of about one-sixth of the households receive a "long form" during the census, which asks for many additional details about the dwelling and its occupants, including income and labor force status.

All this information is used in the redesign to re-stratify the primary sampling units, so as to reflect changes which occur between the censuses. The economic characteristics of many metropolitan areas have changed in recent decades, and there has been a shift of population to some formerly less developed parts of the "Sun-belt" in the southern portion of the United States. Areas which were similar 10 to 20 years ago may be very different today. Taking these changes into account results in more efficient and reliable samples.

Prior to the 1980 redesign, all the surveys used the same general purpose sample design. This general purpose design was modified somewhat for the Current Population Survey (CPS) to improve the measurement of labor force data. The other surveys had smaller sample sizes than the CPS, but used the same PSUs, the same cluster size, and the same within-PSU stratification. Indeed, the other surveys were merely allocated a portion of the extra "reserve" sample which was selected for CPS along with its regular sample. As the importance of the other surveys grew, greater attention was paid to their sample designs. Following the 1980 census, the surveys were redesigned individually, in an attempt to optimize them based on their separate objectives, rather than using a common design modified only for measuring labor force characteristics.

The expenditure and housing surveys sort units based on census characteristics which are highly correlated with the variables measured by the survey. The expenditure surveys concentrate on rent or value of housing, which is asked of all units in the census. The housing survey takes a subsample of census "long form" households, for

which detailed housing characteristics are available. The other five surveys do not sort individual households using census characteristics, either because the relevant questions are not asked in the census, or because the relevant variables are not stable over time and the benefits of sorting would quickly dissipate. Another reason for not sorting separately for all surveys is that some surveys interview clusters of adjacent households to reduce travel costs. Some sorting, such as separating urban and rural areas within each county, is used for all the surveys.

The CPS is still our largest survey and as such still has some effect on the others. The CPS is the only survey that attempts to measure data for states as well as for the United States as a whole. In 1980, the CPS sample was selected as 51 independent state samples, one from each of the 50 states and the District of Columbia. This was necessary because there was a reliability requirement for the unemployment estimates for each state. The need for reliable state data was in response to the allocation of Federal funds determined in part by the estimated state unemployment rates. The other surveys use sample designs aimed at making national estimates, and therefore their primary sampling units may cross state lines.

Although each survey now has its own stratification of primary sampling units, steps were taken in the 1980 redesign to maximize the selection of common sample areas across surveys. This will be done again in the 1990 redesign. This allows field representatives to be shared, and allows the permit and area samples to be better coordinated. The largest metropolitan areas are automatically in sample for all the surveys. Several of the surveys select their sample PSUs as a subset of the CPS PSUs. The crimes survey used a "maximum overlap" method, in which each PSU was given the appropriate unconditional probability of selection, while the expected amount of overlap with the CPS was maximized. (This maximum overlap sample selection is still commonly called "keyfitzing" after Nathan Keyfitz, who developed one of the original methods). A different mathematical method is now used, but the objectives are the same as Keyfitz addressed. The same maximizing technique will be used to select the new 1990 CPS PSUs from their new strata while maximizing the expected overlap with the old CPS sample PSUs. This was done to reduce the need to train new interviewers in some areas while having to lay off interviewers in old areas.

Effect On New Technologies

The last 10 years have seen increasing automation of census data products and survey interviewing techniques. This affects the sample redesign directly because automation offers potential new efficiencies in the sampling operations, and indirectly because changes in design may be needed to make the most efficient use of the new interviewing technologies.

The most dramatic change in interviewing techniques has been the introduction of computer-assisted interviewing, combined with increased use of telephones by the interviewers in the field. The Census Bureau has re-

cently opened a centralized computer-assisted telephone interviewing (CATI) facility, located in Hagerstown, Maryland. Interviewers at the facility work from a computer terminal which automatically selects the next case for interview, schedules callbacks, and displays the questionnaire on the terminal's screen. The computer program determines the path through the questionnaire based on the answers which are entered, checking the responses for consistency as it goes. We expect this system to improve the control of data quality, both because of the control provided by the computer and because interviewers can be closely monitored by the supervisors at the centralized facility. The system eliminates labor-intensive data entry and some steps in processing, which are needed for paper questionnaires.

CATI has been tested successfully for the National Crime Survey and is being tested for the CPS. We expect the CATI methodology to be in full-scale use in the early 1990's. Its use for household surveys has so far been restricted mainly to follow up interviews for panel surveys. An address sample is still used and the first visit to a household is made in person. Even when full CATI is being used, there will still be a need for field interviewers. Not every housing unit has a telephone and some that do request a face-to-face interview. Testing the use of computer-assisted personal interviewing to complement CATI is now underway and is also well underway.

Some testing using samples based on randomly selecting telephone numbers has been done by the Census Bureau. However, the response rates in these tests were much lower than when the first visit was made in person. Because of this, along with the need to represent households without telephones, a purely telephone sampling approach is no longer considered for most of the household surveys. An exception is the Current Point of Purchase Survey (CPP), which is just completing a test of a combined telephone-list/random-digit-dialing approach, with promising results. This could eventually remove CPP from the list of surveys using the census as a frame. We plan to select some CPP sample in the 1990 redesign as a backup strategy in the event that the random-digit dialing approach proves unsuccessful.

One aspect of the 1990 sample redesign will be to modify the sample designs to make more efficient use of centralized CATI. CATI removes some of the follow up interviewing from the dispersed field representatives. This means that the field interviewers will be underutilized unless they are given greater workloads initially. Thus, for a constant total budget, the optimal design using centralized CATI will have fewer PSUs, with a larger initial workload in each PSU. The increased use of telephoning also reduces the relative importance of travel costs, which may reduce the optimal cluster size for those surveys which select clusters of households.

The increased use of computers in the 1990 decennial census will make it easier than ever before to use census data in constructing a sample frame. The most important innovation is a geographic database known as the Topologically Integrated Geographic Encoding and

Referencing (TIGER) system. This computerized system (developed jointly by the Census Bureau and the U.S. Geological Service) will produce a map of any city block or comparable rural "block," with roads and natural features correctly represented. Addresses from the census will be linked to the correct block on the map and in most instances the location within the block will be indicated. The TIGER maps have the potential to revolutionize the area sampling operations. In the past the development of maps has been a particular problem and staff members have struggled with maps and information of inconsistent quality from a variety of sources. The computer generated maps will also simplify locating those new units whose building permits have been selected. The TIGER maps and data (without detailed census address information) will be available to the general public and will be useful for area sampling by survey organizations outside the Census Bureau.

Another 1990 census product which will facilitate using the census as a sampling frame is the automated Address Control File. This contains a record for each census address, with basic information about the housing unit at that address. For about 95 percent of the records, the actual address will be included as text on the file. In previous censuses, the addresses could only be obtained by going to the handwritten register completed by the census enumerator, which necessitated an expensive address keying operation before the surveys could use these addresses.

Computer technology was also used to advantage in implementing the mathematical methods for stratifying and selecting PSUs in the 1980 redesign. Similar methods will be used in 1990. The CPS strata before the 1980 redesign were formed by writing key PSU characteristics on 3x5 index cards and grouping the cards manually to form intuitively homogeneous strata of roughly equal stratum population. For the 1980 redesign, a multivariable clustering algorithm was modified to form strata so as to minimize a measure of total variance for a set of specified variables, subject to constraints on the stratum population. Also, for the 1980 redesign, an improved "maximum overlap" method was developed, which selected a probability sample of PSUs while maximizing the overlap with some other survey's selected areas. This method used a linear programming algorithm to maximize the expected overlap, subject to constraints on the probabilities. This gave a greater percentage of common PSUs than methods used previously.

Coordinating Sampling With Different Sample Designs

A central theme of the 1990 redesign research is to better coordinate the sample selection operations for the different surveys. As I have described already, in the 1980 redesign we allowed different surveys to use different PSUs and different ways of sorting, stratifying, and selecting households within the PSUs. At the same time, every effort was made to keep the surveys' sample units close together to save on the cost of keying ad-

resses, listing for the area sample, and sampling building permits.

This task of linking different sample designs turned out to be quite complicated. An example is the coordination of listings for the area sample. If surveys are to share each other's lists of households in sample blocks, then it is necessary to keep a cross-referenced index so that each survey can find out who else has previously made a list of the block and where that list is being kept. This sort of thing was much easier in the 1970 redesign where there was only the one CPS design, so that it was only necessary to find out whether the block had been listed for the previous CPS sample.

Extensive record-keeping is also needed to avoid duplicate selection of the same address by different surveys. United States Government statistical policy, as set down by the Office of Management and Budget, is that a single address should not be included in more than one Census Bureau survey. With different surveys selecting sample from the same universe, using different method, it was not easy to avoid such duplication. Particular problems were in the Health Interview Survey, which used an area sample where the other surveys were using list sampling, and the housing survey, which selected specific long-form units where the other surveys were using area sampling.

All this complex record-keeping and cross-referencing is amply justified by the large savings from coordinating the survey operations. However, the complexity becomes a liability if at any time between redesigns, one of the surveys has its sample reduced, expanded, or has a change in the scheduled interview dates. When such changes are made, all the references to the changed survey anywhere in the reference system must be checked and updated, to avoid duplication and other operational problems for the other surveys. Because the system was not designed with updating in mind, this causes even small changes in a survey's sample to be time-consuming and expensive, even when the changes can be made by computer.

The sampling of building permits is especially inflexible. Permits are sampled every month as they are issued by the permit offices. (There are over 10,000 of these offices throughout the country.) Many of the offices destroy their old records after a few years, so it is impossible to go back and select more permits.

To try to simplify the record-keeping in the 1990 redesign, we will closely examine the details of all the clerical and computer procedures, and standardize these procedures whenever possible. Some of the major research issues concern whether specific surveys would incur significantly higher field costs or higher variances by simplifying on certain operational details in order to standardize their procedures with the other surveys. In the 1980 redesign, four separate sets of computer programs were used to select the sample for the seven surveys. (Some surveys were able to share programs). These programs were developed separately and there were minor differences in definitions and data formats

which turned out to be major barriers to coordination. In the 1990 redesign, the plan is to use one integrated set of computer programs for the entire sampling operation. In developing these programs, our programmers propose to use a Computer-Assisted Software Engineering approach. Besides providing a common logical framework for the computer algorithms, the computer-assisted planning produces a common data dictionary for all the programs. This will enforce standardization of definitions across the surveys.

Selecting several surveys from the same frame can add complications to the mathematics of sample selection. A basic example is that if a survey removes units from the universe with probability proportional to size, then the remaining universe tends to under-represent large units. Such concerns need to be kept in mind as the sample selection methods are designed. A goal of the 1990 redesign is to leave a "clean" universe, so that future surveys can be selected from the census address lists, building permit offices, and the area frame without having to make special adjustments because of the sample of units which has been "removed" by the redesigned household surveys.

One final challenge in coordinating sample selection for multiple surveys is getting agreement on a time for selecting the sample. It is most economical to do the bulk of the sample selection work at the same time for all

surveys. However, this requires all the different sponsoring organizations to complete their research on the new sample design in time. Some agencies prefer to have their redesigned sample introduced as soon as possible after the 1990 census, to take advantage of the new design. Others would benefit by having more time to use the 1990 census data in research on special topics affecting their survey, before deciding how to design their survey.

"As soon as possible" after the 1990 census turns out to be nearly four years later; the 1990 redesign sample will start being introduced in April 1994. Part of this lag is due to the census processing; the last 1990 census data file used in sample selection becomes available about 18 months after the official April 1, 1990 census day. Once the design has been specified, computer processing to prepare materials for the clerical work takes about 12 months, and the various clerical activities and related processing take about 9 months. This leaves about 9 months to use the 1990 census data to specify the design, including selecting PSUs, deciding how to stratify units within PSUs, and deciding on the sample size at each stage of selection. Obviously most of the basic research, planning, and software design has to take place prior to the availability of the 1990 census data. □

¹Presented at the IFDO/IASSIST 89 Conference held in Jerusalem, Israel, May 15-18, 1989.

NOTICE TO IASSIST READERS

To all IASSIST members

We are trying to collect as many photos as possible taken at IASSIST conferences or other official functions. If you have pictures please send a copy to:

Sue Gavrel

129 Blackburn Ave
Ottawa, Canada K1N 8A6

We hope to have an album or two for the
Edmonton Conference

From Census To Integrated Population Data Or Socio-Demographic Accounts

by Mathieu Vlieghe

Netherlands Central Bureau of Statistics

Introduction

The census of population taken in the Netherlands in 1971 seems to be the last one in a series which started in 1830. In 1981 the government postponed the population census which according to the 1970 Census Law had to be taken in that year. Meanwhile, the government has announced to Parliament a proposal to revoke the 1970 Census Law. At the same time the government submitted an alternative statistical programme consisting of a set of register-based enumerations in combination with survey research during a period of circa ten years. First reactions indicate that Parliament will be in favour of such revocation.

In several publications attention has been paid to some underlying factors with regard to this development (Redfern 1986, Choldin 1987). Relevant for the topic under study here - post-censal surveys - are the public and parliamentary discussions on the privacy issue in relation to a census, in which pleas in favour of an absolute anonymity at taking a census are a central topic: data collection should take place without names and addresses. In this way - it was the argument - the privacy of the individual citizen would be guaranteed. Furthermore, it is worthwhile reminding the demands in these discussions for a voluntary participation in the census by the individual citizen. Every legal obligation in respect to this participation should be rejected, in particular a legally imposed penalty for not cooperating. Such obligations - the argument was - would infringe the fundamental rights of the individual citizen with respect to his willingness to give information about himself to others.

Ultimately those pleas and demands were not yet effective on the 1971 census as such. However, they affected the possibility laid down in the executive regulations of the 1971 census with regard to the keeping of a 10% sample from that census. By linking this sample to the 1981 census it was intended to obtain longitudinal data on among others occupational and educational mobility as well as on changes in household status. To carry out this longitudinal study it should be necessary to keep the names and addresses of the sampled persons during a period of more than ten years.

More and more, however, this procedure was considered in public opinion and in Parliament as an encroachment

on the personal privacy of the citizen. Although the importance of such longitudinal research for statistical purposes was recognized by Parliament, ultimately it was stated that this kind of research connected with a census had to be subordinated to the interests of the individual citizen. Consequently, a few weeks before census date, the Minister politically in charge of the 1971 census was obliged to cancel a possible keeping a 10% sample from the 1971 executive regulations.

Since the 1971 census various sources and methods have been come into use for the construction of a system of population statistics as comprehensive as possible from a demographic as well as from a social and economic point of view (Vlieghe and van de Stadt, 1988). The following developments should be mentioned specifically:

- starting enumerations from the municipal population registers and their respective enlargements in content;
- more extensive exploitation of the municipal reporting with regard to vital events and changes of residence;
- conducting regularly large-scale sample surveys on the labour and housing market;
- application of methods to obtain estimates at the level of the total population;
- the accomplishment of an automated (and yearly updated) register of all addresses in the Netherlands with geocodes, a so-called Geographic Base File (GBF), which - jointly with the municipal population registers - also is in use as sampling frame.

These developments have brought a statistical programme into practice which - in view of the results produced - is highly comparable to that of a conventional census in combination with post-censal surveys. Some parts of this programme show characteristics analogous to a conventional census. In particular this regards the system of demographic statistics which is based on periodical enumerations from the population registers in combination with the processing of monthly municipal reports on vital events and migrations.

To a certain extent this applies also to the large-scale sample surveys which generate benchmark data in the

social and economic field. These surveys have to fulfill this function, since the content of the municipal population registers is restricted to purely demographic data. Of course, in delivering the social and economic benchmarks they cannot completely compete with a census. For example, they cannot provide these data with the same regional detail as a census does. However, the extent of regional detail is sufficient for the level of which national policies regarding the labour and housing market are made.

At the same time, these large-scale sample surveys show characteristics inherent to so-called post-censal surveys. First of all, they provide in-depth information on some specific groups of the population. Secondly, the results of the system of demographic statistics (comparable to the results of a census in the classical sense) are used for making the relevant estimations on the level of the total population from the survey results. Finally, the sources for compiling the demographic statistics (the municipal population registers) themselves are sometimes used as frame for the selection of the sampling units.

These points will be discussed more deeply in the next sections. It should be pointed out, however, that in the system of population statistics built up until now, no use has been made of the technique of record linkage. In the near future the application of this technique will presumably not be used either. Reasons of public and political nature - rather than technical impossibilities - prevent such applications at present. Therefore, the above sketched statistical programme cannot be qualified as a register-based census (Redfern, 1986). Rather it is an approach in which both kind of statistical instruments are jointly applied in such a way that (a) societal changes relative to the field of inquiry in question can be discerned almost continuously by the statistical information provided, and (b) actual developments can be taken as a subject of inquiry in the research programme almost instantly.

The system of demographic statistics

2.1. The continuous population accounting: the basis of the system

The main source for compiling demographic statistics are the municipal population registers. These registers have been introduced in 1850 and set up using the data collected at the population census of 1849. Since then these registers have been continuously updated according to the regulations laid down in the system of population accounting (van den Brekel, 1977). Until World War II the registers consisted of family documents in which all members of the family were listed. Since then the personal card has been introduced. This personal card is made out at birth and follows the individual person during his whole life time.

An essential feature of the population registration system is its decentralization. This implies that each municipality keeps its own population register. Persons are registered in the population register of the municipality

in which they normally reside.

The regulations for the systematical updating of the municipal population registers refer among others to the registration of all changes by birth, death, marriage and dissolution of marriage by the local Registrar of the civil registration to whom they have to be reported. Furthermore they consist of detailed rules with respect to taking permanent residence in and removals from a municipality as well as all changes of residence within a municipality (see figure 1). All these regulations are intended to guarantee the completeness and accuracy of the municipal population registers.

2.2. System of demographic statistics and register-based enumerations.

The system of population accounting also includes regulations concerning the municipal reporting of the various vital events and changes of address to the Netherlands Central Bureau of Statistics (see also figure 1). This reporting enables the CBS to compile continuously statistics on natality, mortality and nuptiality as well as statistics on internal and external migration.

The municipal information on vital events and migrations is also used at the CBS for updating a statistical file with aggregated data on the demographic composition of the population. This file - set up for the first time using the 1947 census data - enables the CBS to compile annually statistics on the size and composition of the population for every municipality.

The file with data on the demographic composition of the population has to be revised regularly. In the course of time deviations from the real situation are inevitably introduced due to the above mentioned method of updating this file. Consequently, the relevant statistics are becoming less reliable over time.

The revision of the 1947 file took place in 1960, still based on the results of the census taken in that year. Since 1971, however, complete enumerations from the municipal population registers are used for revising purposes. The underlying factors for this switch were among others satisfactory results from checks on the quality of the data in the municipal population registers obtained at the 1971 census as well as technological developments in data processing.

At every enumeration the amount of characteristics in the file has been augmented (see figure 2). At present by means of this file statistical information is supplied annually for each municipality on e.g. the total population by age, sex and marital status, and the alien population by age, sex and marital status.

2.3 The system of population statistics: its reliability

The basic demographic data in the municipal population registers have a high degree of accuracy. It is in the interest of the citizen that his data have been accurately recorded in the municipal population register in view of, among others, a request regarding a resident permit, a

driving licence, a passport, as well as several benefits or grants. Moreover, it is in the interest of the municipalities that the data of its citizen are accurately registered. The financial contribution of the central government to the municipalities depends on, among others, the number of its inhabitants. By law, these figures have to be determined annually by the CBS after mutual control with the municipalities. The procedure has been laid down in the above mentioned system of population accounting.

Therefore, the reliability of the demographic statistics is in general high. This can also be concluded from the results of the register-based enumerations carried out for revising purposes (for figures see table 1). The deviations found from the confrontation of the results from these enumerations with those from the yearly updated statistical file are usually very small for various age groups. They are higher for categories of marital status and for the alien population, preponderantly due to incompleteness in the reporting of the relevant changes to the CBS.

2.4. Register-based demographic statistics: some prospects

In addition to the data used in register-based enumerations for revising purposes, the municipal population registers contain still other data which from a statistical point of view are of great value. Up till now these data were not a topic for regular register-based enumerations due to the lack of proper automated processing systems with regard to these registers. Until recently, some municipalities had set up duplicates of their registers in an automated form, but the systems developed were for a great part different from each other. Other municipalities had duplicates of their population register in a mechanised form: either punch-cards or address-plates. Still other municipalities (mostly the smaller ones) had no duplicates at all.

At present an automated system of municipal population accounting - called the Municipal Administration of the Population (MAP) - is developed under the supervision of the Ministry of Home Affairs. Not only the municipal population registers itself is taken into regard, but also the reporting of the relevant changes between the municipalities mutually and between a municipality and its clients (including the CBS). This reporting will take place by means of an electronic network. The implementation of the whole system has been planned to take place within a few years from now.

It is obvious that the automation of the municipal population register in a similar way as described by the MAP will give rise to further developments on the statistical field (Verhoeve and van de Kaa, 1987). For example, by means of a register-based enumeration on the population by status in the family (spouse/alone parent, child, not living in a nuclear family) it will also be possible to compile regularly benchmark statistics on families and (groups of) persons not living in a family nucleus. Already in 1987 such statistics have been compiled. From an organizational and financial point of

view this enumeration had to be restricted to municipalities with an automated register². Moreover, analyses from the results with respect to the number of 'family units' (i.e. nuclear families and persons not living in a nuclear family) living at one address have shown that by means of such a register-based enumeration it is also possible to compile statistics on households, provided that complementary statistical data from other courses (e.g. survey research) are available.

Decisions on the organization of such additional register-based enumerations have to be taken yet. These are dependent on the definite form the municipal reporting to the CBS on vital events and migration will take in the new system of population accounting. In consultation with the Ministry of Home Affairs several alternatives are discussed at the moment.

Finally, the implementation of the MAP will also offer possibilities for improving and enlarging the current demographic statistics. First, individual demographic events could be linked, so that statistics on life-cycles could be compiled. Secondly, demographic statistics could be presented on the territorial subdivision of municipalities formerly used in censuses.

Large-scale surveys

3.1 Introduction

Since the seventies two large-scale sample surveys have been conducted periodically by the CBS: from 1975 the Labour Force Survey (LFS) and from 1977 the Housing Demand Survey (HDS). Both surveys aim at describing regularly the situation at a specific field of inquiry (the labour market, the households and the housing market respectively) as completely as possible, as well as monitoring the developments which are taking place at those fields over time.

Consequently, the statistical information provided by these surveys includes both data which can be used to provide some benchmarks on the relative field in question and so-called in-depth data on those fields. These data are obtained by grossing up the survey results to the level of the total population, the annually compiled demographic statistics being the basis. The principal characteristics of both surveys are described below.

3.2. The Housing Demand Survey (HDS)

3.2.1. Topics: benchmark data and in-depth data

Since 1977 the Housing Demand Survey (HDS) is conducted every four years, partly at the request of the Ministry of Housing, Planning and the Environment. The principal aim of the HDS is to provide statistical information on the present housing situation of the population, the expenditures of the population for housing, realised residential moves in the two years following the survey data (Everaers, 1987a).

By means of this survey benchmark-data can be provided on (a) households, and (b) the dwelling stock and other housing units. Benchmark-data on households concern characteristics such as their size, type and demographic

composition. Benchmark-data supplied with regard to the dwelling stock are type of dwelling, period of construction, number of rooms and type of ownership amongst others.

In-depth data relate mainly to characteristics of households which are of importance for the statistical description of the housing situation. In this respect special attention is paid to ownership or tenancy and the expenditures of the households on housing. Other topics on which in-depth data are provided, are residential moves and potential households, i.e., persons in private households of 18 years and over and personnel living in institutional households who want to move to another dwelling or another housing unit.

The benchmark data are only available for regions with circa 100 000 inhabitants or more. For the housing policies of the government this regional detail is sufficient, since mainly big cities and so-called housing market areas are the target areas in these policies.

As already has been pointed out, in the near future more regional detail in the benchmark data on families (and perhaps on households) will be obtained by enumerations from the municipal population registers. Furthermore, annual statistics are compiled on the size of the stock of dwellings at the level of the municipality. These statistics are based on a yearly updated statistical file set up at the 1971 census. In the coming years this file will be regauged by building up an automated register of dwellings by address.

3.2.2. Sampling procedures

At present two sampling frames are available: the decentralized municipal population registers and the Geographic Base File (GBF). The GBF - a joint project of the Postal Service, the Central Bureau of Statistics and the Government Physical Planning Service - is an automated (and yearly updated) register containing all addresses in the Netherlands including codes for the postal district, the grid square (500 by 500 meters) and the territorial subdivision of municipalities formerly used in censuses.

Theoretically the address together with its known occupants as sampling unit would be the best representation of the target populations of the HDS (living quarters, private households and potential households). In this respect, however, both sampling frames have disadvantages. At present it is not possible to draw such a sample from the decentralized municipal population registers, due to organizational and budgetary problems. Moreover, addresses with vacant dwellings are not included in the population registers. On the other hand, the GBF contains only an indication of the number of postal deliveries and type of building for each address.

Weighting the disadvantages of both sampling frames in connection with the target populations of the HDS, the municipal population registers have been chosen as sampling frame and the person as sampling unit. The disadvantage of having no information on vacant dwellings counterbalances strongly the disadvantages of

having a very high underrepresentation of sub-tenant households (particularly one-person households). Such an underrepresentation was discovered from analyses of results of the 1971 census with those of surveys carried out around 1971 with the address as sampling unit. However, using this sampling procedure the probability of being selected into the sample is not necessarily the same for all households and dwellings. This probability is twice as high for a dwelling occupied by a household with a spouse as the corresponding probability for a dwelling occupied by a household without a spouse. This is a consequence of the research-design: questions regarding occupied dwellings are only to be answered by the main occupant or his (married or unmarried) spouse; questions regarding the composition of households only by the reference person or his spouse. Therefore, corrections are made afterwards for a great deal of the survey-data on more-person households and dwellings (see the next section).

Under the given budget, the sample fraction (1:150) is chosen in such a way that statistics can be compiled with sufficient precision for the big cities and the housing market areas.

The sample is drawn using a two-step method. In the first step a selection of municipalities is made by the CBS. The criteria used are the size of the intended sample fraction and requirements of the fieldwork. In the second step the selected municipalities draw a sample of persons aged 18 years or older from their population register according to written instructions by the CBS. Until now these instructions have to be made separately for municipalities with automated and mechanized data processing systems as well as for those municipalities which still administer their register by hand. Names and addresses of the selected persons (and some registered characteristics) are sent to GBS.

In spite of these instructions there are always some problems in getting samples of sufficient quality from the different municipalities. Some municipalities which have no automated system are unable to draw the sample. Sometimes preselection of persons has been taken place in drawing the sample. Many of these problems can only be solved adequately, when all municipal population registers will be automated.

3.2.3. Method of estimation

The grossed up HDS estimates are based on weighted observations. The respective weights are determined by using the method of post-stratification (i.e. stratification after selection of the sample). In this method the survey population is partitioned into a number of subpopulations, called strata, and all selected persons within a stratum are given the same weight. Per stratum the weight is calculated as the ratio of the size of the total population which is partitioned in the same way and the number of selected persons in the survey (Bethlehem, 1987).

By applying this method both the non-response bias can be reduced and the precision of the estimates at the level

of the total population can be improved. It is known that these effects are only obtained if a relationship exists between the target variables of the survey and the variables used to construct the strata. The data available for constructing the strata satisfy this condition to a great extent.

The respective weights are determined according to the following procedure (Everaers, 1978b). First, weights are calculated for correcting the over- and underrepresentation of population categories in various areas due to selectivity in non-response. The relevant strata for these areas are obtained using the information on a number of registered characteristics of all selected persons received in the sampling stage from the municipalities, such as sex, year of birth, marital status and family status. The selection of the areas is primarily based on the urban/rural distinction.

Secondly, weights are calculated indicating for the various areas the number of persons the selected person represents. The partitioning in strata is based on the municipal information on all selected persons and the results of the demographic statistics on age and marital status. The areas used in this reweighting are the areas for which data of the HDS are published.

Finally, the definitive weights are determined, first, by multiplying the above two weights and, then, by dividing the obtained results by two in those cases where the probability of being selected was twice as high (see section 3.2.2).

3.2.4. Main results and their reliability

The main results of the last HDS are presented in table 2. Their reliability can be checked by comparing these estimates with results from other statistics. Such comparisons can be made with respect to the estimated figures of occupied dwellings and households.

The estimate of occupied dwellings can be compared with the corresponding figure to be derived from two sources, namely: the already mentioned updated 1971 file on the stock of dwellings and the regularly published figures on vacant dwellings. It is found that the HDS estimate significantly deviates from the last figure. However, it has been already noted that the updated 1971 file will be regauged. There are indications that the information on some of the changes in this stock sent monthly by municipalities to the CBS, is unreliable.

HDS estimates on households can be compared with similar information derived from the Labour Force Surveys which up to 1985 have been held every second year. Such a comparison shows a significant difference in the estimated number of one-person households between the two surveys. It is not clear yet which figure could be considered as more reliable.

An indirect check on the reliability of the household estimates can be performed by comparing the HDS population estimates calculated by means of the frequency distribution of the household-size with the corresponding figures from the population statistics. In

table 3 the several figures are given for the total population and for age groups. Looking at this figures one may conclude that the HDS estimates on households on this point seem to be reliable.

3.3. The Labour Force Survey (LFS)

3.3.1. Topics: benchmark data and in-depth data

From 1975 until 1985 the Labour Force Survey (LFS) has been regularly conducted every two years; since January 1987 continuously. Data collection and data processing in the Continuous Labour Force Survey (CLFS) are completely automated (van Bastelaer, 1987). From the beginning these surveys have been designed to provide statistical information on the labour force, educational attainment and qualifications of the population and commuting of the currently active population.

The benchmark data which can be supplied by the LFS refer to among others the main categories of engagement (such as economically active, educational training, engagement in household duties) and not-engagement (such as retirement or disablement); the size and socio-demographic composition of the labour force (including educational attainment) as well as some economic characteristics of the employed persons such as occupation, branch or economic activity, status in employment and place of work. Due to the sample fraction (circa 2.5%) these data can only be published for the administrative areas of the Regional Labour Exchange (64 regions) as the lowest level of regional detail. However, this regional level suffices for national policy purposes with regard to the labour market.

In-depth data are compiled for the employed labour force (for example on several aspects of the time worked as well as of retirement of working; secondary occupation), and for the unemployed labour force (for example on job seeking, registration at a Regional Employment Exchange and social security benefits received).

Moreover, in due time flow data can be provided, since the Continuous Labour Force Sample collects data on the labour history in the preceding year for the population of 15 years or older. This regards among others the dates employment started or ended in the previous twelve months; the main characteristics of the jobs performed during this period such as occupation and branch of economic activity; the reason of terminating a job as well as job seeking activities for every period of unemployment in the previous twelve months. As yet, this kind of data are collected by means of retrospective questions. At present plans are worked out to use a panel for it.

Finally, in the near future further in-depth data will be collected on various additional topics according to a rotating system. Every year a specific topic will be chosen on which monthly information will be collected. At the moment plans are worked out for collecting data on not regular education and training next year.

In the long run the new design of the Labour Force Survey offers the possibility to provide a complementary

set of stock and flow data on behalf of which better insights in the dynamics of the labour market can be obtained. For the present annual figures are published, whilst the compilation of three month moving averages is worked on.

3.3.2. Sample procedures

The Geographic Base File is used as sampling frame and, therefore, the address as sampling unit. Organizational factors and budgetary reasons prevent the use of the municipal population registers as sampling frame, although the person as sampling unit fits the target population the best.

At present households living at ca. 12 000 addresses are visited monthly. This number is halved in the holiday season. In view of requirements of the fieldwork a stratified multistage sample is used. The first stage consists of a monthly revolving sample of municipalities stratified in ca. 80 geographical areas. This stratification is applied in order to obtain reliable annual figures at the levels of the relevant territorial sub-divisions (i.e. areas of the Regional Labour Exchange and areas covered by the regional subdivisions used by the European Communities). The revolving system is applied to municipalities with less than ca. 20 000 inhabitants only and is chosen in such a way that the territorial distribution of the sample over the whole year is as adequate as possible. The municipalities with more than this number of inhabitants are drawn every month.

In the second stage addresses in the selected municipalities are systematically selected. In drawing the sample a double selection probability is given to addresses with more than one "postal delivery". This procedure is applied in order to reduce eventual cluster-effects, since households living at the same address are expected to resemble each other. Therefore, at addresses with a single delivery all households are interviewed; at addresses with more than one postal delivery only half of the households are interviewed. Addresses of institutional households are excluded.

It should be noted that only 4% of the addresses are addresses with more than one postal delivery. The greater part of these addresses regards addresses at which more than one household is living in a dwelling or another housing unit. For the lesser part it concerns addresses with two or more dwellings: not surprising, after all, since the municipalities are recommended to address each dwelling separately.

3.3.3. Estimation method

The grossed up LFS estimates are likewise obtained by assigning weights to the observations using the method of post-stratification. In the biennial surveys roughly the same procedure has been applied as the one mentioned in the section on the HDS. The calculation of weights for correcting non-response effects was based on information from the respondents and - for the non-response - on information from the municipalities sent to the CBS in

connection with the fieldwork which was carried out by municipal civil servants.

For estimating annual figures from the CLFS the weighting procedure used in the biennial survey had to be adjusted. In the adjusted procedure the continuous character of the survey had to be taken into account, in particular the halving of the number of observations in the holiday season. Furthermore, the CBS does not have the relevant municipal information for correcting non-response effects at its disposal any more. Therefore, the number of steps in the revised weighting procedure has been extended.

The definitive weights used for grossing up the sample results are calculated as the product of five intermediary weights. First, a weight dependent on the monthly probability of being included in the sample is given. The second and third intermediary weights are calculated for correcting non-response effects. The second for correcting seasonal differences in the non-response; the third for differences in the non-response by various population categories. In calculating the correction weights for non-response in step two and three the same territorial subdivision is used; the population categories are determined by a combination of the characteristics sex, age and nationality. The partitioning in strata for these areas is based both on the characteristics of respondents and the corresponding demographic statistics.

The calculation of the weights in the fourth and fifth step is intended to get estimates which are representative for detailed population categories (fourth step) as well as for geographical areas on a detailed level (fifth step). In both calculations the same characteristics (sex, age and marital status) in determining the population categories are used, whilst one combination of those characteristics is reducible to the other. This principle of reducibility also applies to the geographical subdivisions used in both calculations. The calculation of both weights takes place simultaneously by iteratively proportional fitting. The strata for the various areas are obtained by using information both from the respondents and the system of demographic statistics.

The estimates are calculated as averages for the whole year. The averages on the level of the total population for the various areas necessary to perform the calculations are obtained by linear extrapolation of demographic figures on the first of January of the relevant year. The extrapolation is based on the demographic developments during the preceding year. The method of extrapolation is applied, since the annual results of the CLFS ought to be published only a few weeks after the fieldwork in December has been finished.

3.3.4. Reliability of results

The reliability of the estimates from the Labour Force Survey - as far as they relate to persons in employment - can be checked by comparing these estimates with data on employed persons which are regularly obtained from (partly integral) surveys among private enterprises and public services. It should be noted that the last-men-

tioned data refer to jobs; the Labour Force Survey, however, to persons having a job. Moreover, in the LFS estimates data on the armed forces and persons employed in households are included; in the results of the establishment-based surveys they are not.

Taking these differences into account the main results of both kind of statistics did not significantly deviate from each other during the period 1975 to 1985. This situation changed at the introduction of the Continuous Labour Force Survey. In comparison with the LFS 1985 the results of the CLFS 1987 show a higher increase in the number of persons employed than could be expected from the increase over this period derived from the establishment-based statistics. This extraordinary increase is but exclusively concentrated under part-time workers with less than 20 hours worked a week. Probably changes in the wording of the questions on employment and a better probing of the CBS-interviewers have led to these results.

Toward integrated population data or socio-demographic accounts

4.1. Separate collection of various benchmark data and coherency in statistical information on the population.

The preceding sections have shown that demographic, social and socio-economic characteristics of the population are collected in connection with the statistical description of a specific field of research and policy. This proceeding has the advantage that coherent statistics can be provided on certain benchmark data and in-depth data on a distinct field simultaneously. In applying this procedure it turns out that for the greater part the data are not tuned to each other. When data obtained in one field are also collected (usually as background information) in another field, very often the relevant figures differ from each other.

Incoherencies in statistical information on subpopulations also exist between results from the above-mentioned large-scale sample surveys and data regularly collected from surveys among private enterprises or institutions of public services. Some examples of the last kind of data are: the data on employed persons already mentioned in the last section, enrolment data obtained from educational establishments as well as data on persons in institutional households based on various surveys among e.g. health care institutions, homes for the aged and other social welfare institutions. Some examples of such incoherencies are given in tables 4 and 5.

The incoherencies are considered unsatisfactory by users of statistical information on the socio-demographic situation of the population. In order to meet the demand for more coherent information on this field the CBS recently started the compilation of Socio-Demographic Accounts (Koesoebjono, 1987).

The underlying aim in compiling these accounts is to provide a coherent statistical description of the socio-demographic composition of the total population in a twofold way. Firstly, on the level of stock data, reflecting size and structure of the population at a certain

moment in time; secondly on the level of flow data, expressing changes in the size and structure of the population between two moments in time.

Achieving data coherency in these accounts necessarily implies a process of adjustments in existing data and of additional estimates for lacking data. A coherent system of stock and flow data requires one and the same reference period, uniformity in concepts and operationalizations as well as an identical target population, i.e. the total population of the country. In this respect it should be mentioned that the existing data (a) relate for the most part to different observations periods, (b) are often based on different operationalizations of concepts and sometimes even on conceptual differences, (c) show differences due to the application of sampling procedures (precision of sampling results, possible sampling errors) and of different estimation methods, and (d) refer to different population categories.

4.2. Methodology of the integration: mean features

As yet the stock data in the accounts refer to the situation at the first of January of each year; the flow data to the period between the first of January of two successive years. The accounts are presented in a matrix form: the stock data in the distributions of the marginal distributions relate, therefore, to the beginning, respectively the end of the period under review; the flow data to the transitions between the categories in these distributions. As a consequence, intermediate transitions are not taken into account.

In compiling the accounts the basic principle in the population accounting is followed: that is, the size of the population at the beginning of a period plus the number of persons entering the populations in the course of the period equals the size of the population at the end of the period plus the number of persons who left the population in the course of the period. This rule is consequently applied for each category which has been distinguished in the matrix.

The process of data integration occurs in various steps. First, the stock data (the marginal totals in the matrix) are compiled. For this purpose quantitative analyses are carried out with respect to the differences mentioned earlier in available data, and adjustments in data as well as minor additional estimates are made. In compiling the stock data the various figures are arranged in order of reliability. In all matrices the population figures are treated as the most reliable ones.

Second, the flow data (the cells in the matrix) are established analogously to the compilation of the stock data. However, in this step more estimates have to be made. Not all data are available, and if they are, they are not directly related to the categories used in the stock data.

In the next step stock and flow data are confronted with each other in the matrix. Explanations for differences and contradictions between stock and flow data are sought for. Thereafter, the relevant figures (on flow and even on stock data) are revised.

Finally, a procedure of iterative proportional fitting is applied in order to obtain a matrix which is internally consistent (that is: the basic accounting principle is valid for each category of the matrix) and which deviates as less as possible from the original matrix (established after the third step). During this process, the stock data are assumed to be fixed, only the flow data change.

4.3. Matrix construction and main results

At present two kinds of socio-demographic accounts have been compiled. One consists of coherent statistical data on the population with reference to type of engagement or non-engagement, the other one with reference to its status in household. The basic matrices - for men and women separately - are very detailed since they contain the data by engagement (status in the household respectively) and age group. The data on sex and age composition of the total population is the framework whereupon the data on type of engagement or not-engagement, and the data on household status are gauged.

Therefore, the first step in the matrix construction consists of the compilation of the demographic data matrix by sex and age. Table 6 shows an aggregation of the demographic matrix for the year 1984. Following the compilation of that matrix, the definitive matrix can be compiled step by step for each demographic population category (by sex and age group). The final result is a matrix by age and type of engagement, respectively household status for men and women separately. An aggregation of the first mentioned data matrix is given in table 7.

Both matrices reflects the composition of the population at two successive moments in time, as well as changes which take place between these two moments. Consequently, the destination of persons belonging to a certain category can be traced at the end of the period. Furthermore, the origin of persons belonging to a certain category at the end of a period can be derived. In this way the respective flows - the outflow and inflow - of each category can easily be calculated. This also applies to a calculation of the turnover flow for each category, that is the numbers flowing into and out of a certain category.

From this point of view the data in the respective matrices can serve as a basis for projections, as among others the destination percentages can, with due reserve, be considered as probabilities of transitions. The availability of a series of such figures over time allows to formulate hypotheses about the future developments with respect to processes of change, and in connexion with this, a projection of the various categories.

4.4. Some prospects

At present integrated stock data with regard to type of engagement or non-engagement are already available for five successive years (1980-1985); integrated flow data for three years. Further developments are directed towards (a) the extension with other categories of type of engagement (e.g. engagement in household duties or in voluntary work) and relevant categories of non-engage-

ment (e.g. retirement, disablement); (b) the construction of such matrices for specific population categories, e.g. the alien population; and (c) the compilation of quarterly accounts in connexion with the development in compiling stock and flow statistics on employment and non-employment based on the CLFS.

The matrices with regard to household status will soon be available provisionally for only one year (1985), due to the lack of relevant data at present. In particular annually compiled statistics on households analogue to the population statistics are missing. Therefore, work is underway to compile such statistics for the short term using different sources, especially the demographic statistics and the CLFS. On the long-term it is expected that household statistics can be compiled regularly by register-based enumerations on family status together with survey data.

Finally, studies are progressing with regard to the presentation of transition within the population in order to have a better insight on its mobility.

Concluding remarks

In the preceding sections a broad outline of the statistical system in the Netherlands has been given as far as this system contains elements in reference to the general topic of post-censal surveys. It has been pointed out that various research and statistical techniques are jointly applied to obtain the relevant demographic, social and economic data on the total population. The instruments for compiling the basic demographic statistics, i.e. the system of population accounting (including the municipal reporting to the CBS) and enumerations from the municipal population registers have been described. Next to this special attention has been given to some methodological aspects regarding the large-scale surveys as being the relevant sources for providing data on the social and economic situation of the population. Finally, a description has been presented on the first efforts to generate coherent statistical information on the level of the total population by means of the development of socio-demographic accounts. This broad outline is summarized in figure 3.

The system of population statistics, the main features of which have been presented above, may be considered as an alternative statistical programme to a population census and post-censal surveys. However, it should be emphasized that this system is not a substitution thereof in the sense that it aims at obtaining exactly the same statistical information. On the contrary, it is to be considered as a procedure of bringing up-to-date the formerly used instruments within existing possibilities and limits posed by (the Dutch) society. Within these possibilities and limits, the systems aims at producing the statistical information users in general are looking for.

References

Bastelaer van, A.M.L., 1987, *The continuous Labour Force Survey*. Netherlands Official Statistics, vol. 2, no. 4, pp. 30-32.

Bethlehem, J.G., 1987, *Weighting sample survey data*. Netherlands Official Statistics, vol. 2, no. 1, pp. 17-18.

Brekel van den, J.C., 1977, *The use of the Netherlands system of continuous population accounting for the population statistics*, Netherlands Central Bureau of Statistics, Voorburg/Heerlen, The Netherlands.

Choldin, H.M., 1987, *Statisticians' responses to the privacy issue*. Paper presented at the meeting of the American Statistical Association, Chicago.

Everaers, P., 1987a, *The Housing Demand Survey 1985/1986*. Netherlands Official Statistics, vol. 2, no. 4, pp. 42-46.

Everaers, P., 1987b, *The pitfalls of secondary data analysis with special reference to the Dutch Housing Demand Survey*. Paper presented at the fifth European Colloquium of Theoretical and Quantitative Geography. Bardonecchia (Italy).

Kosoebjono, S., 1987, *Socio-demographic accounts: framework to measure the dynamics of population*. Netherlands Central Bureau of Statistics, Voorburg/Heerlen, The Netherlands.

Redfern, P., 1986, *Which countries will follow the Scandinavian lead in taking a register-based census of population?* Journal of Official Statistics, vol. 2, no. 4, pp. 415-424.

Redfern, P., 1987, *A study of the future of the Census of population: alternative approaches*. Statistical Office of the European Communities, Luxembourg.

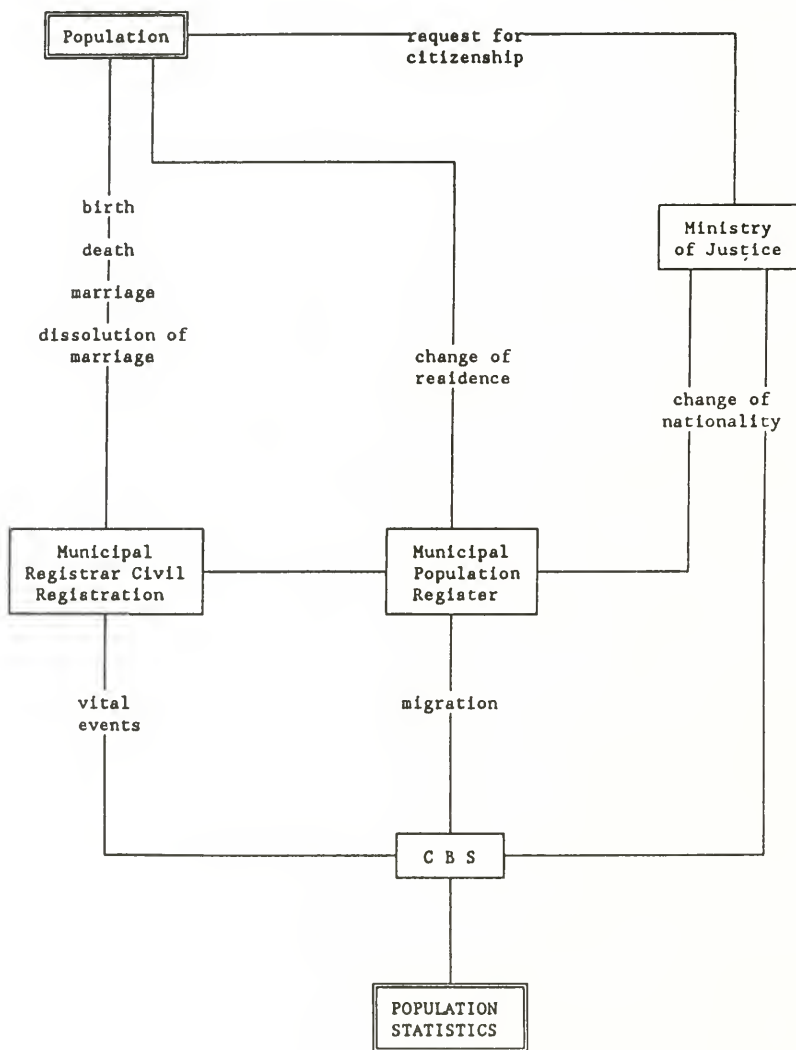
Verhoef, R., and van de Kaa, 1987, *Population registers and population statistics*. Population index, 53 (4), pp. 633-642.

Vliegen, M., and H. van de Stadt, 1988. *Is a Census still necessary? Experiences and alternatives*. Netherlands Official Statistics, vol. 3, no. 4, pp. 27-34.

¹ Presented at the IFDO/IASSIST 89 Conference held in Jerusalem, Israel, May 15-18, 1989. The author expresses his acknowledgements to Santo Kosoebjono for his valuable comments on an earlier draft of this paper.

² Notwithstanding this restriction (nearly 75% of the total population has been enumerated), results have been presented for every municipality by generalizing the results obtained from the automated municipalities to the non-automated ones using their respective composition of the population by age, sex and marital status as a base. Given their geographical position and degree of urbanization, the family composition within the automated municipality was supposed to be equal to the family composition within the similar non-automated municipality.

Figure 1 : From reporting of the population to statistics on the population



**Figure 2 : Synopsis of (planned) enumerations from the municipal registers
for revision purposes, 1971-1990**

1971 : total population by sex,
year of birth and marital status

(yearly
updated)

1976 : alien population by sex,
year of birth, marital status,
and country of nationality

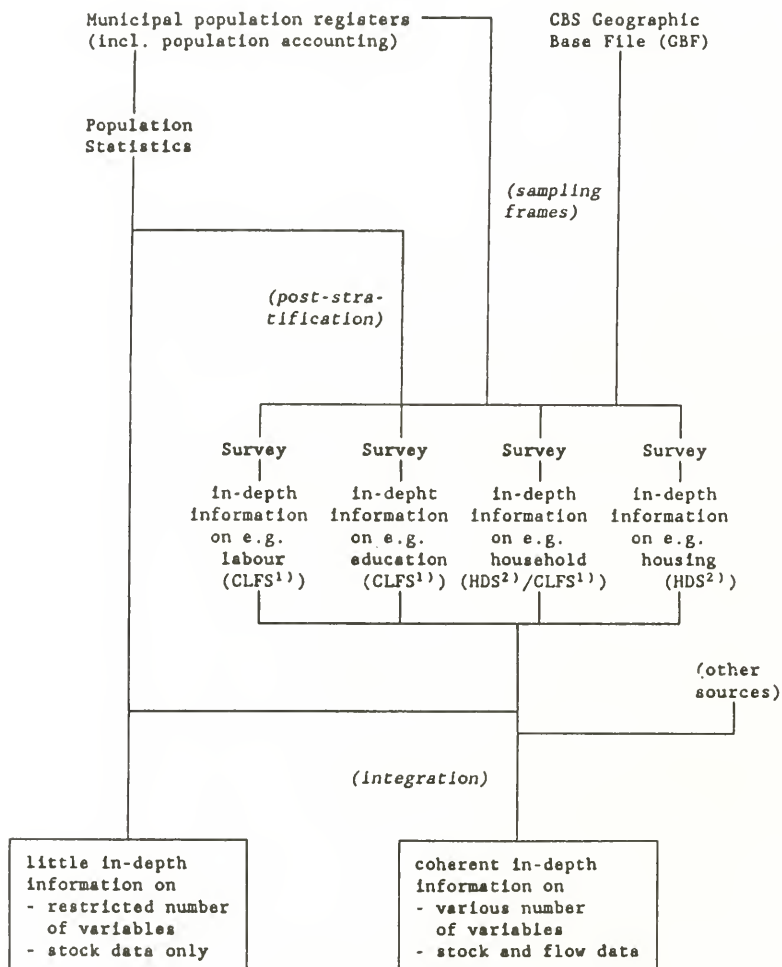
(yearly updated)

1983 : total population by sex,
year of birth, marital status,
and country of nationality

(yearly updated)

1990 : total population by sex,
year of birth, marital status,
country of nationality and
country of birth

Figure 3. From restricted to in-depth information on the population



1) Continuous Labour Force Survey

2) Housing Demand Survey

Table 1. Differences between the results of the registered-based enumeration 1983 and the updated 1971 and 1976 file as a percentage of the relevant categories from the updated files

	Age				Total
	0 - 19	20 - 49	50 - 84	85 years	
	years	years	years	or older	
	I				
Male	-0,0	0,0	0,1	-0,2	-
Female	0,0	0,0	0,1	-0,2	-
Total	0,0	0,0	0,1	-0,2	-
	Marital status				
	never	married	widowed	divorced	
	married				
	I				
Male	-0,6	1,0	-1,2	-2,6	-
Female	-0,3	0,3	-0,2	-0,5	-
Total	-0,6	0,6	-0,4	-1,4	-
	Nationality				
	dutch	alien			
	I				
Male	-0,1	1,6			-
Female	-0,0	0,9			-
Total	-0,1	1,3			-

Table 2. Population, households and housing situation, BGS 1985/1986

	Total	Occupied	Other	In-
	total	dwellings	living	stitu-
			quarters	tions
x 1000				
Households				
one-person households	1 530,7	1 269,8	167,4	
multi-person households	4 034,5	3 993,6	28,2	
total	5 565,2	5 283,4	195,0	
Number of persons	14 402,1	13 908,8	254,2	251,3 1)

1) Persons of 16 years and older

Table 3. Population estimates by age, HDS 1985/1986 and demographic statistics by age, 1986

	To- tal	Age < 15 years	15-29 years	30-49 years	50-84 years	65 years or older
	x 1000					
Demographic statistics	14529,4	2788,2	3725,0	4107,1	2140,0	1769,2
HDS: population in						
-private households	14240,8	2821,1	3657,0	4052,4	2127,9	1582,5
-institut.households 1)	251,3	-	13,0	20,1	13,7	204,4
total	14492,1	2821,1	3670,0	4072,5	2141,8	1786,9
Difference with regard to the demogr. statist.	I - 0,3	+ 1,2	- 1,5	- 0,8	+ 0,1	+ 1,0

1) Population 16 years and older

Table 4. Population in private households by status in household, LFS 1985 and HDS 1985/1986

	One- person household	Multi-person household			
		reference person	spouse 2)	child	other person
x 1 000					
LFS 1985 1)	1 434	4 009	3 589	2 018	202
HDS 1985	1 531	4 035	3 622	4 655	199

1) Population of 15 years and older

2) Incl. living in consensual union

Table 5. Population of 15 years and older in full-time education
by sex and age, ES 1984 (september) 1) and LFS 1985 (april) 2)

	To- tal	Age 15 - 24 years	25 years or older
x 1 000			
ES 1984 (september)			
Male	672	621	51
Female	543	515	28
Total	1 215	1 136	79
LFS 1985 (april)			
Male	632	598	34
Female	571	524	47
Total	1 203	1 122	81

1) Educational Statistics

2) Labour Force Survey

Table 6. Total population by age, data matrix 1984/'85

	stock	of which on 1-1-1985					
	on	in population				not in population	
	1-1-1984	to- tal	0-14 years	15-64 years	65 years or older	death	emi- gration
		x 1 000					
stock on 1-1-1985		14 454	2 850	9 873	1 730		
of which on 1-1-1984							
in population							
total	14 395	14 222	2 861	9 832	1 728	118	55
0-14 years	2 930	2 914	2 681	252		2	15
15-64 years	9 756	9 891		9 580	112	26	39
65 years or older	1 708	1 617			1 617	91	1
not in population							
birth		173	173				
immigration		60	16	42	1		

Table 7. Total population by type of engagement/non-engagement, data matrix 1984/'85 (provisional figures)

	stock	of which on 1-1-1985						
	on	in population					not in population	
	1-1-1984	to- tal	pre- school	full time education	full-time employment	other- wise	death	emi- gration
x 1 000								
stock on 1-1-1985		14 454	700	3 378	4 645	5 731		
of which on 1-1-1984								
in population								
total	14 395	14 222	522	3 359	4 627	5 714	118	55
pre-school	709	704	522	180		2	1	4
f.t. education	3 421	3 405		3 162	138	104	1	15
f.t. employment	4 582	4 547		5	4 273	268	15	21
otherwise	5 682	5 565		11	215	5 339	102	14
not in population								
birth		173	173					
immigration		60	5	20	18	17		

The Impact of Microcomputer Technologies on Dissemination of Integrated Social, Demographic and Related Statistics

by Robert Johnston¹ and
Graham Templeman

Introduction - The Relation Between Integration and Dissemination

The United Nations has been concerned with general issues in measuring development and levels of living and related social, economic and environment conditions since the beginning of the organization, pursuant to the promotion of "higher standards of living, full employment, and conditions of economic and social progress and development" as set forth in the Charter of the United Nations (Article 55). Over the past five years this work has received new stimulus in the United Nations Statistical Office from the great interest at national and international levels in statistics and indicators on women and special population groups such as youth, elderly and disabled persons, and from new interest at national and international levels in the compilation and use of social statistics and indicators to design and implement social policies, to monitor the achievement of social objectives and to monitor the social impact of economic adjustment policies.

In response to these emerging interests and priorities, the Statistical Office has undertaken a substantial reorganization of its methods of compilation, presentation and organization of social and related demographic and economic statistics and indicators. This work follows up the development in the 1970s of the United Nations framework for integration of social, demographic and related statistics (FSDS), and the preliminary guidelines on social indicators published by the United Nations in 1978.² It has been particularly oriented to compilation and dissemination of integrated indicators drawing on a wide range of sources and aimed at non-specialists, rather than compilation of primary data, which are detailed, technical, and specialized by field. This work has been greatly facilitated by the rapid development and now nearly universal availability of highly standardized microcomputer hardware and software technologies for statistical work.

As described in the United Nations Handbook on Social Indicators³, the development of integrated social statistics and indicators is a wide-ranging and multi-faceted process which aims to bring together basic statistics from many different fields and data collection programmes and recompile them for many different

purposes. The Handbook provides a basic core of structure, concepts and methods for use in this process and thereby promote the compilation and dissemination of social statistics and indicators to better meet a wide variety of user needs through more effective integration and use of the basic data.

Unfortunately, the cost and methodological difficulties of bringing together social statistics and indicators from many disparate and often intractable primary and secondary sources have held back work on social indicators in many countries and internationally. Even where statistical services have built up a considerable volume of basic data, this has by no means ensured the ready availability to users of indicators relevant to their specific purposes and concerns, including policy issues. For such purposes, close collaboration between users and producers of indicators, and detailed attention to data requirements for indicators at the stage of designing data collection programmes, and co-ordination within a framework such as FSDS are needed. Thus much of the work on social indicators in the 1970s and early 1980s was concerned with precise identification of user interests and requirements and their translation into well-structured statistical methods and concepts.

Co-Ordination and Integration of Social and Related Classifications for Dissemination of Integrated Statistics and Indicators

One of the technical features which works on FSDS and has been stressed from its inception has been the development and harmonization of social and related classifications for integration and for indicators. However, the development and harmonization of social and related classifications is a complex process for many of the same reasons that social statistics and indicators present statistical offices with so many difficult problems of organization and methodology. In general, the subject matter is extremely heterogeneous and the relevant statistics come from a very wide range of sources, each with established traditions, procedures and objectives and often administered more or less independently of the central statistical service. These circumstances are similar at national and international levels.

The initial development of FSDS in Towards a System of Demographic and Social Statistics and in the preliminary guidelines on social indicators established a basic

subject-matter and classifications framework for the further development of classifications and indicators. These early reports served to clarify what classifications were relevant to integrated social statistics and in what ways. With the current interest in improved multidisciplinary compilation and dissemination of social statistics and indicators and given the new technical possibilities of microcomputers for bringing together data in micro-computer data bases, the importance of harmonized classifications emerges all the more clearly.

A basic principle of work in the Statistical Office in this area continues to be the importance of the close linkage between so-called basic statistics and statistics for integration, and for indicators that are, basically for general disseminations. Thus it has never been suggested that new classifications should be developed for integration or for indicators, which would in any case be a technically and organizationally impossible task, given the degree or decentralization of responsibility for statistical classifications at national and international levels and the large number of competing interests and technical problems that must always be delicately balanced in preparing any kind of recommendation on classifications. What the Statistical Office undertook in the preliminary guidelines on social indicators and has now been made much more explicit in the Handbook of Social Indicators, is to recommend abstracting from existing classifications shorter forms which are needed for integration and for indicators. As the draft Handbook states, once the fields and topics for indicators have been outlined in an indicators programme at the national or international level, basic statistical classifications for use in indicators should be developed. These must, of necessity, be based on the classifications used in the basic data but for purposes of indicator compilation these source classifications often require careful adaptation.

The process of adaptation should be undertaken with three objectives in mind:

- (a) Meeting specific indicator requirements;
- (b) Abbreviating classifications as much as possible to simplify compilation and presentation of indicators;
- (c) Devising classifications into which data from a variety of sources often using differing classifications or variants of classifications, can be fitted as consistently as possible;
- (d) Identifying population groups of special policy concerns.

All of the classifications referred to in the illustrative series and basic data tables for indicators in the Handbook are listed in the table below which also shows the fields in which they are used. Sixteen of these are considered basic classifications in the Handbook. Five of these concern demographic and social characteristics (sex and age group, national or ethnic group, household size and composition, household headship and level of education); three are geographical (urban and rural areas, cities and urban agglomerations, and geographical

regions); four concern activity characteristics (occupation, status in employment, socioeconomic group and time-use); and four are classifications from economic statistics (percentage distributions of household income and consumption, kind of economic activity (industry), functions of government and institutional sector).

These basic classifications can be used to provide a firm foundation for the development of indicators in all of the fields covered by the Handbook. They were selected for discussion as basic classifications on the basis of (a) their substantive importance for indicators, usually in more than one field and drawing on multiple data sources, (b) the extent of their importance and use for indicators in national and international experience, and (c) the relative detail and complexity required in their use for compiling statistics for indicators. All but one of the basic classifications are shown and discussed in the illustrative formats for basic data tables of the Handbook, drawing on the relevant international recommendations. The exception is classification by national or ethnic groups. In this case, national experience and circumstances are so diverse that no international recommendations are feasible and even an illustrative classification could not serve any useful purpose.

Principles of Integration Applied to Dissemination of Statistics and Indicators on Women and Special Population Groups

Interest in the development of statistics and indicators on women and other population groups that are considered to be of special relevance for policy planning has given considerable impetus to a range of activities concerned with statistics and indicators on these groups. The principal groups on which work has been concentrated in the United Nations Statistical Office are women (beginning with the World Conference of the International Women's Year in 1975), disabled persons (beginning with the International Year of Disabled Persons in 1982), youth (in connection with International Youth Year in 1985) and children. There has been interest in the development of statistics and indicators on the elderly (in connection with the World Assembly on Aging in 1982 and the International Plan of Action).

In international compilation and dissemination of indicators on women, for example, a substantial quantity of data is being routinely collected in international statistical services and supplemented, in many cases, with standardized international estimates and projections. The rapid spread of microcomputers and the ease of use of spreadsheet techniques have now made it feasible to compile these data in one source, using the FSDS framework, disseminate them to users cheaply and quickly on diskettes, and prepare user-oriented software and documentation for reference, analysis, table-generation and similar uses. A special project with these objectives was established in the Statistical Office in 1984, and this work was basically completed in 1987.

The United Nations Women's Indicators and Statistics Data Base (WISTAT) consists of 72 microcomputer

spreadsheet files (currently using Lotus 1-2-3) ranging in size from approximately 20kb to 150kb and totalling about 12mb. WISTAT is available from the Statistical Office on 22 microcomputer diskettes complete for 178 countries and areas or for specific regions, using the forms provided in the printed user's guide (currently available, in part, as a Statistical Office working paper). WISTAT will be fully documented in the user's guide, to be issued in final form as a sales publication of the United Nations. A listing of statistical series and topics in this data base is given in the annex below.

Using quite different underlying technical methods of organization and compilation but identical microcomputer hardware and software an international disability statistics data base was also completed by the Statistical Office in 1987, comprising detailed statistics on disabled persons from censuses and surveys in 55 countries and areas between 1975 and 1985. Like the women's data base, the Disability Statistics Data Base (DISTAT) is disseminated on diskettes. It consists of 34 microcomputer spreadsheet files ranging in size from about 7kb to 314kb and totalling about 3.3 mb. The files are described in detail in United Nations Disability Statistics Data Base, 1975-1986: Technical Manual.⁴ <footnote text> The complete data base is available from the Statistical Office on 12 microcomputer diskettes using the forms provided with the Technical Manual. Version 1 of the data base (as of 31 December 1987) contains (a) information on sources and availability of statistics on disability for 95 countries or areas for various years between 1960 and 1986, and (b) detailed statistics on disabled persons from national censuses, surveys and other data sources from 55 of those countries or areas for the period 1975-1986.

Finally, the basic strategy and framework for organizing social statistics for social indicators, as set out in the Handbook on Social Indicators, were adopted by the Statistical Office for preparation of the Compendium of Statistics and Indicators on the Situation of Women - 1986 and the Compendium of Social Statistics and Indicators - 1986.⁵ That is, highly simplified basic data were compiled from primary international sources into microcomputer spreadsheets. Once in the spreadsheets, new series and indicators could be calculated and data transferred within and among spreadsheets with great flexibility and minimal time and effort and, once final table formats were agreed, they could be tested and then generated in final form very quickly. On this basis, series and classifications such as those given in the Handbook have been prepared for these two compendiums, with the possibility of recalculating percentages, rates, ratios, distributions, reagggregations and the like and of juxtaposing series from different sources that may be of interest almost at will. A short, preliminary version of the social compendium was prepared using these techniques for the United Nations Interregional Consultation on Social Welfare Policies and Programmes held in September 1987⁶ and generated considerable interest among delegates with no special statistical background.

Overall, it appears that microcomputer hardware and

software for spreadsheets, data bases and analysis are at the leading edge of basic changes in the development of social statistics and indicators at national and international levels. The effects are now beginning to be seen on a wide scale and at the same time the technologies are advancing and spreading so rapidly throughout the world that the direction and full implications of these changes are still not completely understood or fully appreciated.

The Potential Role in Dissemination of Spreadsheets and Data Bases on Microcomputers.

Spreadsheets and their users

Spreadsheets are used extensively by people interested in statistics, and are a useful tool in most cases.

Spreadsheets are characterized by a row/column cellular approach to data in which each cell may be (typically) a number, a character string, or a numeric or logical function of the values in other cells. Cells are named by a column/row coordinate system. For example:

	A	B	C	D ...	AA	AB	AC...
1							
2							
3							
.							
.							
.							
11							
12							
.							
.							
.							

Spreadsheets are capable of displaying or printing the cells in row/column format, and manipulating the cells individually or by rows, columns, or groups of these. They are, effectively, the "cell processing" equivalent of a word processor. A word processor imposes minimal structure on its atomic units (words) except to organize them within given boundaries such as paragraphs and margins. A spreadsheet, on the other hand, maintains a positional relationship between its atomic units (cells) which is capable of being displayed as a two-dimensional row/column table.

The facilities for manipulation offered by spreadsheets and word processors have many parallels, such as "cut and paste", insertion and deletion, and search and replace. Word processors have some facilities peculiar to themselves, such as reformatting between redefined

margins, text flow from line to line, and so on. Spreadsheets also have some peculiar features such as row and column transposition, default cursor movement by row or column first, and so on.

It is not surprising, therefore, that people dealing with statistical tables are drawn to spreadsheets in much the same way that typists and writers are drawn to word processing programs, by the facility for direct interactive visually-based control.

Beyond these "cell-processing" capabilities, some spreadsheets offer what they call "database capability". (See, for example, Lotus 1-2-3 Tutorial Manual, v.2.01, p. 5.1)

Spreadsheets “database capability” comes from an analogy with some aspects of relational database theory. This type of data organization can be simulated using a spreadsheet, by using the relation name as the spreadsheet name, and by treating rows as records or instances of the relation and columns as fields or attributes, as long as the user sets up the data in the format of a single relation.

When the user sets up the data in "flat file" format, i.e. with one column representing the key field, all columns with unique names, and so on, the row and column manipulations available to the spreadsheet user parallel some of the simpler facilities available to a relational database user. For example, sorting rows by the key field, searching a column for a particular value or for values falling within a range.

In some spreadsheets such as Lotus-1-2-3 it is possible to use foreign key values located by a search such as a range search to extract rows from another spreadsheet. In this way the "cut and paste" spreadsheet operations simulate the linking of different relations. When such operations are expressed as a macro, i.e. a named sequence of operations which can be invoked by its name, the result can be quite efficient for some purposes. It is important to realize, however, that with spreadsheets there is only a limited connection between the column name and the data. The true column name is its coordinate. If a column is removed or added to a spreadsheet, thus pushing other columns to new coordinates, any operation which refers to coordinates will have to be redefined. This is not a problem for perfectly stable data. Spreadsheet statistical tables in presentation format. Tables with hierarchies of field names simply do not lend themselves to database-style manipulations other than those which can be simulated by "cut and paste" cell

processing.

A second consequence of the affinity of spreadsheets for human readability is that spreadsheet designers tend to follow the make use of the horizontal left to right direction of reading. This results in the expansion of more variables horizontally than vertically, creating hierarchies of column headings. In the WISTAT data collection, for instance, almost every spreadsheet uses one row per country or area, with column hierarchies up to four levels deep. For example see table below:

Some simple arithmetic shows that this hierarchy generates 24 columns and that the "adult" heading will appear 8 times. An interrogation asking for listing of those countries for which, in 1975, the number of adults

[illegible]

prosecuted exceeded a certain number, would require two of the twenty four columns to be identified, then summed, then compared to the given reference number. The first step, identifying the columns, can only be done by a person or by software capable of dealing with hierarchical field structures. For similar reasons, the task of sorting this particular data by number of persons prosecuted within age-group or sex would be daunting.

In the WISTAT data collection the tendency to horizontal expansion of variables has an even more direct problem. Many indirect users want to extract data for some or all subject areas for one country or region. Country or region is typically the only vertically expanded variable, i.e. it identifies the rows. When the relevant row from each spreadsheet is identified and extracted it is not possible to combine these rows into a new spreadsheet or an integrated listing because each row has a different column structure.

The result of extracting all data for one country is approximately 70 one-line spreadsheets. If titles and column headings are extracted as well, then we have approximately 70 spreadsheets each with a dozen or so rows. Thus the most appropriate application for spread-

sheets is the manipulation of cells and blocks of cells after the table has been formed in the desired way by tabulation or data management software.

Database theory and statistical data

When people talk about a "database" they are usually referring to the type of organization of data which allows the retrieval and display of items of subgroups of the data by name or description rather than by reference to where or how they are stored. Strictly speaking the software which carries out access and retrieval is an essential component of the database.

A typical database query would be "list the names and addresses of all respondents who are hostile to interviewers". Or for aggregated data: "list the countries for which female constitute more than 50 per cent of the population". Queries for an aggregate statistical database would also be expected to extract and format tables, either for printing or for export to spreadsheet files. The particular way of expressing the question depends on the query language defined by the database software in use. One type of database organization is known as "relational". This approach is very popular, having received excellent coverage in computer magazines, and many data management packages claim to be relational. The most well-known aspect of relational data management is that its fundamental data structure is a "table". This makes it attractive to people who like to think of their data in tabular format.

The relational "table" is, however, very different from a statistical table, and in many ways the relational approach is not suited to statistical data.

In fact the fundamental data structure is a relation. A relation is like a pattern. Any type of entity for which data is to be held is given a relation or pattern, which is a list of named place-marker. For example, if we are to hold data about staff we would define a staff relation specifying the items of data to be held:

staff (id-number, name, department, position,
date-hired, type of contract, etc.)

One of these items, or a group of them, must function as a unique identifier, or "key". It is shown underlined here.

The actual data consists of "instances" of the relation, e.g.:

(658889, Templeman, DIESA, 6 JUN
88, consultant, etc.)

When a set of instances is listed it looks like a table. Such a set may be stored in traditional computer terms as a file with a key field and a simple set of fields, sometimes known as a flat file.

When data is to be stored about entities which have a specific relationship to each other, relational theory specifies the mechanism for relating them. For example,

there may also be a Department relation, e.g.:

Dept (Dept-id, dept-name, location, name-of-head, phone-of-head)

There is a specific relationship between staff and departments. This is expressed by including the key of the department relation as an ordinary field of the staff relation. It is known as a "foreign key":

staff (id-number, name, dept-id, location, ...)

For data to function relationally it has to be set up specifically to do so. Data about one entity should not be embedded within the relation for another entity, multiple field values are not permitted, hierarchical field structures are not permitted, and so on. When data comes naturally with such impurities of structure it must first be converted into a logically equivalent set of relations of the acceptable type. This conversion process is known as "normalization". □

¹Presented at the IFDO/IASSIST 89 Conference held in Jerusalem, Israel, May 15-18, 1989.

^{1b}The authors are, respectively, Chief of the Social and Housing Statistics Section of the United Nations Statistical Office and consultant to the Statistical Office on the United Nations Women's Indicators and Statistics Data Base (WISTAT). The views expressed are those of the authors' and not necessarily of the United Nations.

²See Towards a System of Social and Demographic Statistics. Studies in the Integration of Social Statistics: A Technical Report, Improving Social Statistics in Developing Countries: Technical Report, Studies in Methods, Series F, Nos. 18, 24 and 25 (United Nations publications, Sales Nos. E.74.XVII.8, E.79.XVII.4, E.79.XVII.12), and Social Indicators: Preliminary Guidelines and Illustrative Series, Statistical Papers, Series M, No. 63 (United Nations publications, Sales No. E.78.XVII.8).

³Series F, No. 49 (United Nations publications, in press).

⁴Series Y, No. 3 (United Nations publications, Sales No. E.88.XVII.12).

⁵Series K, No. 5 (United Nations publications, in press), and Series K, No. 6 (United Nations publications, in press).

⁶"Compilation of selected statistics and indicators on social policy and development issues" (E/CONF.80/CRP.1).

Comparative Charting of Social Change in Four Industrialized Societies

by Simon Langlois¹

Institut québécois de recherche sur la culture
et Département de sociologie, Université Laval

Contemporary analysts of social change have somehow abandoned the ambitious tenet of building a general theory which could predict or explain all the contemporary forms of social change. A search of laws of history, or works on the stages of development (Rostow), are no more pertinent. R. Boudon has proposed an analysis of the reasons why such a loss of interest for a general theory on social change occurred.

According to Boudon, social systems are not regulated by general laws which are extensible to all societies. Consequently, the search for universal laws must be abandoned and analysis of social change must give up the "monologic" perspective (Boudon 1983:104). In its place, he suggests to draw some possible statements (in French, "énoncés de possibles"), i.e. relationships which have a certain probability, in a certain location and a certain time, instead of general laws of the form of conditional laws or transitional laws. From the Boudon's individualistic approach, there is no more universal laws because sociologist's observed propositions are in part the results of aggregate actions of individuals.

Almost all general propositions (laws) in social sciences are in fact situated observations which are not universal, and it could be almost possible to pose counter-examples. Otherwise, Boudon suggests to consider the existing theories of social change as formal models which need to be adapted to every specific and concrete situations.

Ted Caplos also criticized social change studies, more precisely empirical studies, saying that the parallel between social change and technological change is misleading. "Instead of being continuous, like scientific-technological progress, social change seems to be episodic, non-consultative, non-consistent and non-reversible" (Caplow, 1988:3).

Too severe a diagnosis? In fact, these two briefly referred to comments, from different perspectives, reveal or show a growing uneasiness with the traditional ways of studying social change.

A new approach is essential, which will make possible to observe and to consider the diversity of ongoing social change, and, more precisely, which facilitate the observation of non-cumulative and non-consistent change. This

can be done with social trend analysis.

Social trend analysis

Social trend analysis is sitting between social reporting and system analysis of the global society.

Social indications "designate statistics that are supposed to have significance for the quality of life, and sets of social indicators designate a social report" (Mikalos, 1982:2). There is a normative perspective in constructing social indicators. They intend to measure goal attainment, or to evaluate and compare the respective performance of different countries. By themselves, social indicators do not form a social system. They are built to measure different aspects or dimensions of social life, or to evaluate current public policy and programmes. International comparisons of social indicators are aimed at comparing differential goal attainment from one country to another. Countries are in fact located on a continuum. Consider, for example, life expectancy in good health after 65. Behind this social indicator, there is a clear or precise objective: staying in good health as long as possible. If this indicator increases, one expects to say that the quality of life is increasing; and the quality of life is supposed to be higher in a society in which this indicator is also higher.

"Social indicators facilitate establishing social goals and policies. For example, a time series comparing life expectancy of several countries shows that life expectancy in the United States is lower than in the United Kingdom, Canada, Japan and Sweden. Such information on levels attained by other countries shows us what is attainable and stimulated action" (Ferris, 1988:609-610).

Opposed to social reporting, one finds global diagnosis which summarize, under a single heading or a macro-trend, a great lot of particular trends: post-modern society, dependent society, traditional or modern society, etc. A large number of segments are postulated to go in the same direction, and the macro-trend either summarizes these particular or specific trends, or is presented as the cause which orientate them in a certain direction: this approach raises many questions: to what extent all the trends are really convergent? For example, is France really a post-modern society? Is it possible to speak of the process of Europeanization in contemporary United States (Oxford Analytica)?

Trend analysis, in the perspective proposed here is somewhere between the two approaches stated before.

A French research team, who have published an article under the pseudonym Louis Dim, define trends in this way: The trends which have been identified are sometimes in the nature of behaviour which can be isolated and taken as an indicator of a more general movement. For example, the decrease in religious practice or the increase in participation in sport or sports-related activities. Other trends are much more global, they formulate an overall judgement on a social sector or on an aspect of society. They refer to a sociological theory by which the phenomenon is delimited and can be judged. (Louis Dim, 1985:401)

The analysis of trends is often associated with the study of phenomena which are quantifiable or which have direct relevance to government policies (income, unemployment, population, etc.). Some trends are better documented than others, mainly because they are based on data or indicators gathered by a government statistical office, indicators which have, in a way, received the lion's share of analysis. Changes in income, voting, fecundity, unemployment, for example, are better known than those in division of domestic labour or in use of leisure time. However trends can be analysed other than quantitatively, it is also possible to discern the direction of phenomena with a certain amount of accuracy by using qualitative studies or monographs: strengthening of kinship, increases in new forms of religious practice, etc. It is this broad concept of trends such as these that we are speaking of.

The trend is not an indicator nor a statistical series. It is a diagnosis on a social segment, narrowly defined (declining fertility, for example) or much more wide (increasing mobility of daily life). In the first case, the trend is in fact close to an indicator, but in the second one, it comes from the convergence of many indicators, of many statistical series.

Our unit of analysis is a trend, i.e. a series of values representing the incidence of some item of social behaviour in a given population at points of time in a consecutive sequence. Guidelines, annexed).

The trend is a sector-based diagnosis of the changes in and the direction of a phenomenon or of an aspect of social reality, for example, a drop in the birth rate, an increase in disposable income per inhabitant, an increase in poverty, a decrease of the inequality between the sexes, an increase in the importance of kinship, a growth of individualism, a decrease in the practice of religion, etc.

Trend analyses are in fact studies of the present situation in light of the past. It must not be confused with prediction of the future, nor with futurology. It is a way of studying changes "en cours". This perspective is important for another reason: the comparative purpose.

Theoretical framework and method

The array of trends to be analysed is extremely broad and we did not want to use too narrow a theoretical framework. The choice of trends cannot be firmly fixed since it will be modified somewhat as the analysis goes along, as the French experience has shown. The choice is not based on definite theoretical bases, thus we are not limiting ourselves to the study of marginal trends which may reveal tomorrow's norms, for it is far from certain that this will be the case. The critique of Reich's work, the well-known essay on the new culture published in the 60's, by Hamilton and Wright (1986) is revealing in this respect. It is no longer necessary to limit oneself to known territory, because analysis should also be able to discern new trends, especially outside the sphere of work, although this remains important. This explains why the choice of trends to be studied is based mainly on a group of hypotheses, not a single one, and remains open to change. The trends are in fact derived from sociologically driven categories.

We have agreed with the other research teams to give priority to behaviour, ritualized situations in institutions and to structures, giving lesser importance to values and social perceptions. The priority given to recently identified trends does not mean that the realm of social perceptions will be completely absent. Social representation is less "objective", for lack of a better term, than behavior, at least a priori. People's jobs, salaries, levels of education, whether they have a religious or civil marriage, their actual number of children, are things which can be observed quite accurately, taking errors of measures into account. Aspirations as to salary, the feeling of being deprived, the kind of marriage intended, the number of children desired, satisfaction in a couple's relationship, are all kinds of perceptions and attitudes. Measures of these things are not only filled with errors, they are also more unstable. The number of children, salary or level of education are characteristics which are probably more stable than aspirations, for example.

As to theory, it seems difficult to omit the domain of social imagery and to only take into account behaviour, or more factual data on individuals. Let us not forget that actors also give a meaning to their conduct. Measuring perceptions, while difficult, thus seems relevant and necessary to an in depth understanding of social phenomena and conduct.

The analysis of relations among trends is intended to be still more inductive. At this stage of the project, we want to hypothesize as little as possible about the relationships. Critical analysis of the literature shows that this inductive approach can be very fruitful. We are not starting out with a general theory about global society, such as those of D. Bell, H. Braverman and others, no matter how attractive and pertinent they may be. We intend rather to work out empirically what the overall interpretation might be, somewhat like the method of Oxford Analytica in *America in Perspective*. An example will illustrate our method. It is already possible to gather, from the analysis of several indicators, some

overall trends: an increase in individualism, mobility in daily life, a change from hierarchy to network, etc. These broad trends are the result of observation of a large number of indicators. It is now the analyst's task to interpret the, to discern all the implications, so that the process results in a tentative generalization or theory about all the data.

It must be noted that the proposed study of trends will try to work out, as far as possible, the variations in the sub-groups such as age, sex, social and cultural group and region. The trends are not consistent and enormous differences exist side by side. This has been shown in several recent studies, Bella's work, for example. The analysis of trends will not be limited to the study of average or means only.

Studying trends in four industrialized societies

Our proposed research project intends to identify the principal trends with characterize global society by doing secondary analysis of existing data. Therefore, the first task will be to trace and synthesize published or available observations and analyses.

Up to now, nothing original, for such studies yet exist in number, on all the possible objects. For example, Canadian Social Trends publish excellent analysis on different indicators and different trends. The same for publications like British Social Attitude, Social and Cultural Report (Holland), works published by Eurobarometers, etc. There are fewer analyses of relationship between trends, and fewer again are the attempts to study systematic relationship between a large set of trends. This is precisely what we are planning to do. The level of the proposed analysis is somewhere between the social reporting and the systemic analysis of a large, global, macro-trend which summarize a very large sample of specific trends.

The Louis Dim team in France took the initiative by inviting other research groups from various countries to undertake a comparative analysis of trends, adopting as closely as possible the methods which they have worked out over the last several years. Three teams have already agreed to participate in the project, one from the USA, led by Theodore Caplow, a German team led by Glatzer and Hondrich, and an IQRC team from Quebec, led by S. Langlois. A preliminary meeting was held in Paris in May of 1987 to establish the main goals of the project and to choose which trends which would be observed. The project at hand is now part of a co-operative venture involving several teams from various countries, teams which share the same goal and approach to the study. The Quebec team, based at IQRC, has been given the secretarial role in this little group, coordinating the development of this project.

The area to be covered and the indicators are not chosen for the purpose of verifying a particular theory of social change, nor to illustrate a dominant trend (for example, the increase in individualism in contemporary society). We know as well that many of the indicators of social change which are used, in scientific analyses and in

public debate, reflect an era when the majority of the population spent most of its time working and the times when survival was a daily challenge: unemployment, standard of living based on steady income, poverty, jobs and social position were and are the indicators most relied on in studies of the social structure. We do not deny their relevance, quite the contrary, but it seems to us necessary to develop others which can reveal ongoing social changes (owning a second home, mass media consumption, touristic travel abroad, etc.), indicators which are usually outside the sphere of work which will point out new trends. Anything to do with the world of work will, however, still have an important place in the analysis of trends.

The final choice of the data to be analysed, of the trends to be examined, was based on hypotheses in existing monographs or put forward in some theories about social change, but also, relied on the observations of researchers and on research going on elsewhere, always taking available data into account.

A number of 79 trends were identified at the first meeting of the international group in which our team is participating, based on a preliminary list drawn up by the French team. We agree that the area to be covered is vast. It can be done, however. The Louis Dim group's first attempts which is going on in France showed this. Basically, it is a question of drawing up a brief synthesis of what is known, in the form of trends, about each of the things on the list.

Each team undertakes an analysis of the trends which describe its own society, while following the common method as closely as possible, basically that suggested in the article by L. Dim, 1985, in order to allow for later comparative study. The importance, or the interest, of a comparative study is obvious. It is, among other things, an excellent way to the extend the analysis of causal relations. Let us look at a known example: Has development of education promoted an increase in intergenerational social mobility? There may be a whole new light put on the analysis of relations between these two trends after comparative examination by four societies.

This comparative study of different societies poses enormous problems. Faced with these problems, there are two attitudes to take. We could do nothing, since there are so many difficulties. Or we would try to iron out the problems in order to prepare a trial method, however imperfect it may be at first. This is what we decided upon. In order to smooth out the difficulties, it was agreed that a common grid of theses for the study of trends would be drawn up and that, as much as possible, the representative data concerning the overall society would be used and the measures to be used would be clearly elucidated. But, above all, it was agreed that the proposed comparative analysis would bear on the direction of trends and the relations among the trends themselves. This approach minimizes the problems of comparison to some extent. Thus, it is not a rate or a precise measure which is to be compared (rate of unemployment, real income, etc.) but the direction of a trend and its relation to other trends.

The data

We will be carrying out a secondary analysis of existing data and a synthesis of published works on a given subject. We will be trying to obtain a set of statistics, standardized as much as possible, starting from 1961 if possible, or from 1970, so as to have at least about fifteen years of observation in order to discern a trend. Where statistics may not be available, we will look for data observed for at least three separate periods of time, again so as to discern trends. This will be done mainly for the secondary analysis of data from surveys and the study of changes in social perceptions.

We have already identified about 250 series of statistics with which to characterize the trends we will deal with. Others will be added as the project advances, because the research consists precisely of identifying or even constructing such series for later analysis. The list of data is too long to include here. Examples are number of automobiles, real income per capita, circulation of daily newspapers, etc.

Priority will be given to quantitative data which can be compared in a given society and later between societies. Preliminary examination indicates that for the majority of trends in the attached list it is possible to find out at least one set of statistics. This preliminary data base will be completed by qualitative observations and analyses or by monographs concerning "phenomena portending the future", always working from secondary sources. These observations will clarify the process of some extent.

Scientific and social significance of the project

This project may seem ambitious to some. However, the French experience during the past four or five years proves that it can be done and the results can be fruitful. The scope of the project is broad: the diagnosis of global society and its social changes using secondary analyses. One of the project's interests is to use existing data in order to arrive at a more in-depth analysis.

The comparative dimension should also be stressed. Despite the difficulties that this presents, we believe that the results will be productive. The comparison with other countries will allow us to go ahead with certain interpretations. □

COMPARATIVE CHARTING OF SOCIAL CHANGE

LIST OF TRENDS

(Québec, December 1988)

0. Context

- 0.1 Demographic trends
- 0.2 Macro-economic trends
- 0.3 Macro-technological trends

1. Age Groups

- 1.1 Youth
- 1.2 Elders

2. Microsocial

- 2.1 Self identification
- 2.2 Kinship networks
- 2.3 Community and neighbour-hood types
- 2.4 Local autonomy
- 2.5 Voluntary associations
- 2.6 Sociability networks

3. Women

- 3.1 Female roles
- 3.2 Childbearing
- 3.3 Matrimonial models
- 3.4 Women's employment
- 3.5 Reproductive technologies

4. Labour Market

- 4.1 Unemployment
- 4.2 Skills and occupational levels
- 4.3 Types of employment
- 4.4 Sectors of the labour force
- 4.5 Computerization of work

5. Labour and Management

- 5.1 Structuring of jobs
- 5.2 Personnel administration
- 5.3 Size and types of enter-prises

6. Social Stratification

- 6.1 Occupational status
- 6.2 Social mobility
- 6.3 Economic inequality
- 6.4 Social inequality

7. Social Relations

- 7.1 Conflict
- 7.2 Negotiation
- 7.3 Norms of conduct
- 7.4 Authority
- 7.5 Public opinion

8. State and Service Institutions

- 8.1 Educational system
- 8.2 Health system
- 8.3 Welfare system
- 8.4 Presence of state in society

9. Mobilizing Institutions

- 9.1 Labour unions
- 9.2 Religious institutions
- 9.3 Military forces
- 9.4 Political parties
- 9.5 Mass media

10. Institutionalization of Social Forces

- 10.1 Dispute settlement
- 10.2 Institutionalization of labour unions
- 10.3 Social movements
- 10.4 Interest groups

11. Ideologies

- 11.1 Political differentiation
- 11.2 Confidence in institutions
- 11.3 Economic orientations
- 11.4 Radicalism
- 11.5 Religious beliefs

12. Household resources

- 12.1 Personal and family income
- 12.2 Informal economy
- 12.3 Personal health and wealth

13. Lifestyle

- 13.1 Market goods and services
- 13.2 Mass information
- 13.3 Personal health and disorders beauty practices
- 13.4 Time use
- 13.5 Daily mobility
- 13.6 Household production
- 13.7 Forms of erotic expression
- 13.8 Intoxication

14. Leisure

- 14.1 Amount and use of free time
- 14.2 Vacation patterns

14.3 Athletics and sports

14.4 Cultural activities

15. Educational Attainment

- 15.1 General education
- 15.2 Professional education
- 15.5 Continuing education

16. Exclusionary Phenomena

- 16.1 Immigrants and ethnic minorities
- 16.2 Crime and punishment
- 16.3 Behavioral and emotional
- 16.4 Poverty

17. Attitudes and Values

- 17.1 Satisfaction in life domains and in general
- 17.2 Perceptions of social problems
- 17.3 Orientations to the future
- 17.4 Values
- 17.5 National identity

¹Presented at the IFDO/IASSIST 89 Conference held in Jerusalem, Israel, May 15-18, 1989

The International Research Group for the Comparative Charting of Social Change

(Club de Québec)

Guidelines For New Members

Introduction

The International Research Group on the Comparative Charting of Social Change in Advanced Industrial Societies, informally known as the Club de Québec, was founded in 1986 by a group of sociologists and historians from France, West Germany, Canada, and the United States, who had been studying social trends in the respective countries. These separate studies attracted a fair amount of scholarly and popular attention but did not advance our general understanding of contemporary industrial society as much as they should have, for want of a comparative perspective. Without systematic international comparisons, it is impossible to know whether trends we discover in national societies are local accidents or features of a larger system.

Another reason for combining our efforts was that we had been using different methods to delineate social trends and wanted to standardize them to facilitate comparison.

The project was initially sponsored by Council for European Studies, but each national team has found the financial support of its own research operations. The Canadian team assumed responsibility for maintaining a central secretariat, at the Université Laval in Quebec. Meetings have been hosted in rotation by national teams: at Charlottesville in April 1986, Paris in April 1987, Bad Homburg in June 1988, Quebec in December 1988, and Charlottesville again in May 1989.

At Quebec in December 1988, responding to a request from Professor Constantin Tsoukalas of the National Social Research Center in Athens, the members voted to admit additional national teams on the following conditions:

1. Persons proposing to establish a national team, and professional personnel subsequently added to a national team, must be individually approved by a vote of the general membership of the International Research Group.
2. National teams are obligated to follow the research methods adopted by the International Research Group, including the standard format of trend reports, the list of trends for which reports are prepared, and the criteria for acceptable data.

3. Trend reports may be written in any language, but each national team has the responsibility of eventually making its reports available for publication in English.

4. Members of any national team shall have full access for their own scholarly purposes to the data gathered by other national teams, and shall make their own data reciprocally available.

The Research Problem

Social change is too large a topic to be manageable without further specification. We are specifically interested in the late twentieth century, the industrialized or partly industrialized nations and the social structures and institutional patterns that characterize the behavior of mass societies, especially those associated with the family, voluntary associations, work, leisure, education, religion, government and politics. Our unit of analysis is a trend, i.e. a series of values representing the incidence of some item of social behavior in a given population at points of time in a consecutive sequence. Most of our work has been done with time series 10 to 60 years long, ending as recently as possible, and covering such matters as family income, household expenditures, employment and unemployment, working conditions, the informal economy, marriage and divorce, household composition, kin networks, housing, migration, educational achievement, criminality, leisure patterns, health care, social movements, and so forth.

In the scholarly literature, a few trends have received the lion's share of attention. Economists have looked very closely at trends in economic growth, prices and wages. Political scientists have studied twentieth century trends in voting and party affiliation. Demographers have scrutinized trends in fertility and mortality. It is no coincidence that these are the areas of social life which lend themselves most readily to quantification and offer the longest time series. But a description of social change that limited itself to trends in economic development, political participation and population would be incomplete indeed. Even though quantification is initially more difficult in other institutional sectors, many of the difficulties have been overcome in recent years, and we can anticipate that the quality of data will continue to improve.

The International Research Group's standard list of

trends and indicators currently includes 77 trends grouped into 18 major categories. It is attached hereto. Note that the development of a national profile calls for the preparation of 77 trend reports, each corresponding to one of the numbered subheadings in the list, from 0.1 Demographic Trends to 17.5 National Identity. The first four national teams have already finished this phase of work, which takes from 12 to 36 months depending upon the man- and woman-power available, and they are currently engaged in the more challenging task of comparing their results. We anticipate that new national teams will complete their national profiles at various times during the next five years and we believe that this staggered schedule may be intellectually advantageous.

Following the list of trends and indicators, you will find, first, the standard format for trend reports, and second, a review of the project's long-term goals by two participants.

Format for the Presentation of Trend Reports

Coverage

A trend report is prepared for each numbered subheading in the List of Trends and Indicators, e.g. 0.1 Demographic Trends, 0.2 Macroeconomic Trends, 0.3, Technological Trends. The items listed underneath each subheading constitute a check-list of topics that should be covered in the trend report, but the checklist is not intended to be inclusive.

Arrangement

A trend report normally has four sections: a brief summary of about five lines at the beginning, an explanatory text, a section of tables and figures, and a bibliography. Each table or figure is placed on a separate page. Factual statements in the explanatory text, as well as all tables and figures, should be fully referenced, but the bibliography is usually much more extensive than would be required for direct referencing alone.

If possible, trend reports should be prepared on an IBM or IBM-compatible personal computer, using Word Perfect for the text.

Labels

Every page of every trend report is labeled for each identification, as in this example:

[CCSC-US, 3/8/88, Trend #4.1, Draft #3, AC, Page #4]

Interpreted as follows:

CCSC-US: Comparative Charting of Social Change,

United States

3/8/88: Date Prepared

Trend #4.1: Taken from List of Trends and Indicators,

4.1 Unemployment.

Draft #3: Self-explanatory

AC: Preparer's initials, A. Carrier in this case

Page 4: Of this trend report

Criteria for Data

1. The data were obtained by empirical measurement or observation.
2. The data refer to an entire national society or a representative sample thereof.
3. The data may be expressed as a time series.
4. The time series covers a period of at least ten years, ending in 1983 or later.
5. The time series include measurements or observations for three or more time-intervals, obtained contemporaneously.
6. The data are defined in such a way that the measurements or observations can be replicated in other national societies.
7. The data are defined in such a way that the measurements or observations can be replicated in the same national society in future years.

The Contemporary Jewry Database: Coordination and Flexibility in Bibliographical Registration and Retrieval

by Mira Levine, MLS¹

Director of The Bibliographical Center

Institute of Contemporary Jewry of the Hebrew University Jerusalem

The Bibliographical Center in Contemporary Jewry was designed some five years ago to compile a computerized database of bibliographical information relating to various aspects of 20th century world Jewry. A digital minicomputer (VAX-750) was purchased and the ALEPH (Information Retrieval program developed at the Hebrew University was adapted for this purpose. In the fall of 1985 we began registering indexed and annotated descriptions of books, articles, films, taped interviews and other archival material at or available to the Institute of Contemporary Jewry. The records are accessible both for publication as separate catalogues by the cooperating bodies and for on-line global searching via the ALEPH network of the Hebrew University and other Israeli academic institutions.

To date, over 15,000 bibliographical items have been indexed, abstracted, and registered by seven cooperating but separate computer projects:

Bibliography in Anti-semitism Through the Ages of the Sasson International Center for Research in Anti-semitism

Steven Spielberg Jewish Film Archive - holdings and catalogues

Jewish Filmography Project listing - of films on Jewish subjects available throughout Israel

Oral History Division of the Institute of Contemporary Jewry - holdings and catalogues

Publications of the Institute of Contemporary Jewry, and of its teachers and researchers

Bibliography of Contemporary Jewry 1984-1987, listing relevant books and articles published between 1984 and 1987 arriving at the Jewish National and University Library. (Although discontinued for lack of funds, it will be resumed when funding is available.)

Studies in Contemporary Jewry in English and Hebrew - articles, book reviews and books reviewed in the two annual journals of the Institute of Contemporary Jewry.

Slated for future inclusion are the library and bibliographical projects of Jewish Demography, syllabi of courses offered through the years at the Institute of

Contemporary Jewry, American and the Holy Land bibliographical and archival project, and other funded projects approved by the Institute. Once registered, all records are retrievable via a single master index - by author, title, subject or words appearing in the title, subject headings of abstract. For example, by entering the search term "intermarriage" one summons a list of not only books and articles in which intermarriage is treated, but films and taped interviews as well. Further sophistication in word searching is to be included this summer in the newest ALEPH version, thus facilitating access even more.

Having noted briefly the construction and composition of the Contemporary Jewry Database, let us turn now to the various products and activities it generates at the Institute of Contemporary Jewry Bibliographical Center.

Printed catalogues and bibliographies:

Several participating collections retrieve and publish their respective entries as catalogues, bibliographies or filmographies. The Bibliographical Center staff designs the ALEPH application, advises, trains and supervises staff involved in all aspects of registration and retrieval right up to the preparation of photo-ready copy for the publisher. Garland Publishers in New York has undertaken the publication of sections of the database as part of its series *Garland Reference Library of Social Science*. To date, one volume has appeared:

- *Anti-semitism: An Annotated Bibliography, Volume 1*, edited by Susan Sarah Cohen for the Vidal Sasson International Center for the Study of Anti-semitism, 1987.

- the second volume of the Anti-semitism bibliography

and three others are in the final stages of preparation

- *Films of the Holocaust: An Annotated Filmography of Collections in Israel*, edited by Sheba Skirball for the Spielberg Jewish Film Archive (in preparation)

- *Oral History of Contemporary Jewry: An Annotated Catalogue*, compiled by Institute of Contemporary Jewry.

On-line database

Of course, the primary product of the bibliographical center is an efficient, easily used on-line database of bibliographical information on the various aspects of twentieth century Jewry. Subjects include the Holocaust, Zionism and the State of Israel, Jewish demography and other social science research, Anti-semitism, Israel-Diaspora relations, Jewish communities the world over and the Arab-Israeli conflict to name the most salient.

Thesaurus generation

A third product, an on-line publishable thesaurus for Contemporary Jewry, will be discussed in detail later on.

Network searching

Besides catalogue production, database maintenance and thesaurus development, the Bibliographical Center provides researchers at the Institute with access to information relevant to their work via retrieval from our own database, libraries and databases on the network of Israeli universities (ALEPH), and other Israeli and foreign databases accessible via our facilities. Individual search and subject updating requests are filled as much as time and budget allow.

In coming months we hope to be extending these search services to include access to Israeli databases of relevant materials outside the ALEPH system as well pertinent foreign databases and vendors such as DIALOG and BRS. Institute of Contemporary Jewry researchers are also assisted in computerizing their own research using the ALEPH system.

Experience in cooperative computerization

One of the most interesting by-products of developing the database has been the experience and knowledge that has accumulated in the process of designing and implementing our cooperative computerization. We are frequently turned to by libraries, archives and information centers interested in establishing similar or related projects, or in implementing the ALEPH system for non-library and/or multi-media applications. We enjoy these opportunities to share both our knowledge and data. At the same time we benefit by learning about and often gaining access to related data registered at these institutions. Thus, the spirit of cooperation has been extended beyond the parameters of our own institution's projects.

ALEPH adaptation

We have found the ALEPH information retrieval program particularly suitable for our independent/interdependent applications. First, its multi-lingual, multi-character-type capacity is essential for registration of materials describing world Jewry in Hebrew and Yiddish as well as Latin character languages; Arabic is also available, though not used by us; Cyrillic and Far Eastern character use is still to be developed. ALEPH's structure of several local libraries within a single global library allows, on the one hand, independence in design, cataloging and maintenance, searching and retrieval, and on the other hand, overall maintenance and control if desired.

Other invaluable aspects of the program for our purposes are its user friendly presentation (as we have over a dozen professionals and countless users at varying levels of computer proficiency), and its flexible thesaurus construction and maintenance capabilities. Being part of the overall network of Israeli academic libraries and databases is of great advantage, not only by providing access to our database from any of the 25 installations all over the country, but by allowing us to search their collections as well with but a simple 5-character command. Finally, the excitement generated by participating in the development of such *au courant*, constantly growing and improving research project has stimulated creative applications, fruitful cooperation and productive commitment by our bibliographical center and participating project staffs.

Coordination and flexibility

Indeed, the execution of such a multi-faceted, multi-disciplinary and multi-media database has been a fascinating, albeit challenging, exercise of coordination and flexibility. Each stage of designing, implementing, evaluating, improving and expanding the database over the past five years has necessitated a careful balancing of desire for overall standardization while satisfying the requirements peculiar to each individual project.

Cooperation began with sharing information on and the costs of hardware acquisition and maintenance. Experimenting with ALEPH adaptations for bibliographical and archival applications also involved learning from previous and each other's insights and mistakes. Common code assignments wherever possible facilitates maintenance as well as the sharing process. Meanwhile standardizing divergence provides helpful searching and maintenance cues. For example, all codes peculiar to a given collection are preceded by the same character: P for the Film Archive, A for Anti-semitism, while both PTL and ATL are dumped into and accessible via a single title index.

Training for indexers, abstractors, catalogers, and editing is centralized as much as possible, and retrieval for publication, contacts with both our publisher and computer facilities of the university are centrally coordinated. This avoids wasteful duplication of efforts while at the same time providing supportive, stimulating collegiality. Peculiarities of different types of material, support organizations or lines of authority thus become fruitful bases for comparison and adaptation.

Two activities at the bibliographical center illustrate particularly well the balance between inter-dependence and independence of the cooperating members of the Contemporary Jewry Database: one is thesaurus control, the construction and maintenance of our controlled language for indexing; the other is catalogue preparation, the production of photo-ready copy of bibliographies/catalogues/filmographies for publication.

From its earliest stages, the preparation of photo-ready copy for publication had involved cooperation and individual tailoring. Basic designing of print formats, registration codes and procedures, uploading and downloading, and word-processor refinement are the same. Thus joint sessions for training and exchange of experience are both time-saving and mutually beneficial as are coordinated communications and contractual arrangements with common publishers and sharing of equipment. At the same time, flexibility is required to accommodate the inevitable differences unique to each publication. Filmographies tend to be published in one form, bibliographies in another, and oral history catalogues in still another. Various editors even at the same publishing companies have differing format preferences as do the academic and managing editors of each volume. Thus retrieval and printout design must be sufficiently flexible to satisfy the needs of both producers and users of the catalogues. Needless to say, the experience of producing one publication provides the basis for evaluating, redesigning and/or adapting not only the next in that series but all others emerging from the database. Coordinated quality control to insure continued acceptance of manuscripts thus becomes another important function of the bibliographical center.

Nowhere, however, are the benefits of coordinated interdependence and flexibility-promoted independence more evident than in the construction and maintenance of our controlled language for indexing purposes. At least once, and sometimes twice a month representatives from each of our participant collections meet to discuss specific subject headings for inclusion in our thesaurus. We are joined by those responsible for thesaurus control in such related projects or institutions as the Truman Institute Library, the Mount Scopus Central Library for the Humanities and Social Science Library, the Index to Periodicals in Jewish Studies and the Library of the Melton Center for Jewish Education in the Diaspora. Not only are individual terms decided upon, but all related terms and how these relationships are to be expressed are also discussed.

This year, for example, we have reviewed, among others, our use of the term Holocaust and its various categories, Israel-Palestine-State of Israel as well as our SEEF/SEE/SEEA, BT/NT rubric. Our aim is standardization wherever possible. However, various factors not infrequently require us to modify this goal. Some members are committed by choice or convenience to international thesauri and are not free to adopt all of our joint decisions. Then again, the nature of some material dictates different emphasis. Most of our material is verbally, conceptually oriented, yet film indexing by definition must be visually oriented. Truman Institute's subject approach is geographical - developing countries; thus their library's index terms all begin with the name of the country involved.

Indeed, one might ask how any co-ordination is possible or whether the effort is worthwhile. Certainly the discussions with others facing mutual problems, listening to their proposals for solutions and how they arrived at

this is in itself helpful. But beyond that, we have developed the method, when standardization seems either impossible or undesirable, of at least referring also to the variant uses, so that users searching in any of the collections participating in these thesaurus meetings are made aware that different terminology may also be worth checking. Such co-operation has been most satisfying, particularly for our desire as librarians to make the maximum amount of material available to the maximum range and number of users.

In conclusion, a few general remarks about flexibility and coordination are in order. First, the independent/interdependent approach arises from the very complexity of the project. Second, it is facilitated, indeed made possible by the above-mentioned features of the ALEPH programs as we have implemented it. Finally, a few words about the "humanization" of computerization as opposed to the much discussed and even more feared mechanization of human life.

In no field is there greater resistance to computers than in the humanities. Dedication to the printed word and book in hand, coupled with the complexities of digitizing verbal concepts, not to mention fear of machines, have prevented many humanities teachers and researchers from enjoying the comforts even of simple word-processing. Yet, in our project, as in many others, the cost, complexities and constantly changing nature of the computerization process has, if any thing, increased human contact and cooperation: equipment, programs, staff, knowledge, expertise, data are all enhanced by sharing and mutuality. Thus balancing inter/ and independence is probably the most effective way of developing computerization in humanities - and perhaps human life in the age of computers. □

¹Presented at the IFDO/IASSIST 89 Conference held in Jerusalem, Israel, May 15-18, 1989.

NEWS RELEASE

CENTER FOR MACHINE-READABLE TEXTS IN THE HUMANITIES

Rutgers and Princeton Universities have received grants from the National Endowment for the Humanities (Program for Research Tools, \$30,000), the Andrew W. Mellon Foundation (\$50,000), and the New Jersey Committee for the Humanities (\$10,000) to undertake jointly the planning for a Center for Machine-Readable Texts in the Humanities. Project staff include Marianne Gaunt (Director), J. Penny Small, Kathleen Ciociola (Rutgers); Robert Hollander, Judith Rowe (Princeton); Leslie Hume (Research Libraries Group). Members of the Advisory Board for the planning process are Nancy Ide, Vassar College; Robert Kraft, University of Pennsylvania; Michael Sperberg-McQueen, University of Illinois-Chicago; Donald Walker, Bellcore.

During the course of the planning period project staff will be investigating issues related to the establishment of a cooperative center which will act as a central source of information on humanities datafiles and a selective source of datafiles themselves. The initial goals of the Center as outlined in the project proposal are: the continuation of an on-going inventory of machine-readable texts; the cataloging and dissemination of inventory information to the broader scholarly community; the acquisition, preservation and servicing of textual datafiles which would otherwise become generally unavailable; the distribution of such datafiles in an appropriate manner; and the establishment of a resource center/referral point for information concerning other textual data. Organizational issues, technical issues, intellectual ownership, access and dissemination, and physical facilities are broad areas of investigation during the planning process.

The Center does not propose to duplicate the archives and repositories that already exist for the collection and dissemination of textual data, but rather to complement existing collections and to bring bibliographic control to existing datafiles. To that end project staff will be networking with existing projects/centers to establish appropriate means of collecting inventory data for the cataloging of archival holdings. Progress reports will be sent to publications of scholarly associations and announced on HUMANIST. A listserver has also been set up for communication among those interested in the project and invites inquiry and advice.

Bitnet communications may be sent to: Gaunt@zodiac.rutgers.edu or to BobH@phoenix.princeton.edu. Mail addresses are: Marianne Gaunt, Alexander Library, Rutgers University, College Ave., New Brunswick, N.J. 08903; Prof. Robert Hollander, Dept. of Comparative Literature, Princeton University, Princeton, N.J. 08544.



INTERNATIONAL ASSOCIATION FOR
SOCIAL SCIENCE INFORMATION
SERVICE AND TECHNOLOGY

.....
ASSOCIATION INTERNATIONALE
POUR LES SERVICES ET
TECHNIQUES D'INFORMATION EN
SCIENCES SOCIALES

Membership form

The International Association for Social Science Information Services and Technology (IASSIST) is an international association of individuals who are engaged in the acquisition, processing, maintenance, and distribution of machine readable text and/or numeric social science data. The membership includes information system specialists, data base librarians or administrators, archivists, researchers, programmers, and managers. Their range of interests encompasses hard copy as well as machine readable data.

Paid-up members enjoy voting rights and receive the **IASSIST QUARTERLY**. They also benefit from re-

duced fees for attendance at regional and international conferences sponsored by IASSIST.

Membership fees are:

Regular Membership: \$20.00 per calendar year.

Student Membership: \$10.00 per calendar year.

Institutional subscriptions to the quarterly are available, but do not confer voting rights or other membership benefits.

Institutional Subscription:

\$35.00 per calendar year (includes one volume of the Quarterly)

I would like to become a member of IASSIST. Please see my choice below:

- ☐ \$20 Regular Membership
☐ \$10 Student Membership
☐ \$35 Institutional Membership

My primary interests are:

- ☐ Archive Services/Administration
☐ Data Processing
☐ Data Management
☐ Research Applications
☐ Other (specify) _____

Please make checks payable
to IASSIST and Mail to:

Ms Jackie McGee
Treasurer, IASSIST
% Rand Corporation
1700 Main Street
Santa Monica

Name / title

Institutional Affiliation

Mailing Address

City

Country / zip/ postal code / phone

CURRENT RESEARCH

in library & information science

- **CURRENT RESEARCH** is an international quarterly journal offering a unique current awareness service on research and development work in library and information science, archives, documentation and the information aspects of other fields
- **The journal provides** information about a wide range of projects, from expert systems to local user surveys. FLA and doctoral theses, post-doctoral and research-staff work are included
- **Each entry provides** a complete overview of the project, the personnel involved, duration, funding, references, a brief description and a contact name. Full name and subject indexes are included
- **Other features include** a list of student theses and dissertations and a list of funding bodies. Each quarter, an area of research is highlighted in a short article

CURRENT RESEARCH is available on magnetic tape, as well as hard copy, and can be searched online on File 61 (SF=CR) of DIALOG

Subscription: UK £86.00
Overseas (excluding N. America) £103.00
N. America US\$195.00

Write for a free specimen copy to

Sales Department
Library Association Publishing
7 Ridgmount Street
London WC1E 7AE
Tel: 01 636 7543 x 360

LA

LISA

Library & Information Science Abstracts

- **international scope and unrivalled coverage**
LISA provides English-language abstracts of material in over thirty languages. Its serial coverage is unrivalled; 550 titles from 60 countries are regularly included and new titles are frequently added
- **rapidly expanding service which keeps pace with developments**
LISA is now available monthly to provide a faster-breaking service which keeps the user informed of the rapid changes in this field
- **extensive range of non-serial works**
including British Library Research and Development Department reports, conference proceedings and monographs
- **wide subject span**
from special collections and union catalogues to word processing and videotex, publishing and reprography
- **full name and subject indexes provided in each issue**
abstracts are chain-indexed to facilitate highly specific subject searches
- **available in magnetic tape, conventional hard-copy format, online (Dialog file 61) and now on CD-ROM**
Twelve monthly issues and annual index

Subscription: UK £157.00
Overseas (excluding N. America) £188.00
N. America US\$357.00

Write for a free specimen copy to
Sales Department
Library Association Publishing
7 Ridgmount Street
London WC1E 7AE
Tel: 01 636 7543 x 360

LA



Automation in Archives

The central theme of this meeting is using computers to assist in administering and servicing collections, especially in small institutions.

- Preconference workshops
 - Planning for Automation*
 - USMARC AMC Format*
 - Appraising Electronic Records*
 - Records Management*
- Automation projects tours
- Using PCs to support archival appraisal, description, reference, and institutional management
- Ten Software Packages: Users' and Vendors' Perspectives
- Database Design
- Art and Architecture Thesarus Workshop
- Mixing Archaeology with Automation
- Several Non-automation Sessions

- Vendors' Exhibits

Old Colony Inn, Alexandria, Virginia
Reasonable Room Rates
Free Parking

For further information contact: Don Harrison (202) 501-5575

MARAC Fall 1990 Conference, Alexandria, Virginia
November 1-3, 1990

Please send me: ☐ Preregistration package
☐ Membership application

To: Jim Byers
Archives of American Art
Smithsonian Institution
Washington, DC 20560

Name: _____
Address: _____



LIBRARY

SEP

UNIVERSITY OF NORTH CAROLINA

SPOTLIGHT
ON THE
UNIVERSITY OF NORTH CAROLINA
LIBRARY

FIRST CLASS

SEP 11 1999
CALIF

107
FEB 1999